

版权注意事项：1、书籍版权归著者和出版社所有；
2、本PDF仅用于个人获取知识，进行私底下知识交流；
3、PDF获得者不得在互联网以任何目的进行传播；
如有需要，请尽量购买正版实体书！支持书籍作者！！

大数据金融丛书

区块链 与大数据

打造智能经济

井底望天 武源文 赵国栋 刘文献 ◎ 主编

BLOCK CHAIN
AND BIG DATA



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS

大数据金融丛书

区块链 与大数据

打造智能经济

井底望天 武源文 赵国栋 刘文献 © 主编

BLOCK CHAIN AND BIG DATA

人民邮电出版社

北 京

图书在版编目 (C I P) 数据

区块链与大数据：打造智能经济 / 井底望天等主编

· 一 北京：人民邮电出版社，2017.6

ISBN 978-7-115-45740-0

I. ①区… II. ①井… III. ①电子商务—支付方式—研究 IV. ①F713.361.3

中国版本图书馆CIP数据核字(2017)第089583号

内 容 提 要

近年来，大数据在迅猛发展的同时也充斥着概念的炒作，面临着诸多困境。那么，有什么方法能够促使大数据突破困境，获得更高层次的发展呢？这就是本书的核心思想及价值所在。

《区块链与大数据：打造智能经济》提出了运用区块链技术解决大数据发展中所面临问题的思想。本书分为8章，首先通过对大数据发展所面临困境的分析，引出了迅速崛起的区块链技术，比较了两者的技术沿革及异同；然后通过分析数据流通中存在的问题，揭示了区块链技术带来的改变；最后介绍了区块链技术在相关行业中的应用，并对区块链技术与大数据的结合展开了畅想。总之，本书站在科技发展的前沿，对区块链与大数据的融合发展进行了深入阐述，力求为行业发展提供可行性的指导建议。

本书适合政府、企业、科研机构、金融机构人员以及高等院校相关专业的师生阅读。

◆ 主 编 井底望天 武源文 赵国栋 刘文献

责任编辑 张国才

责任印制 焦志炜

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

大厂聚鑫印刷有限责任公司印刷

◆ 开本：700×1000 1/16

印张：15.5

2017年6月第1版

字数：180千字

2017年6月河北第1次印刷

定价：59.00元

读者服务热线：(010)81055656 印装质量热线：(010)81055316

反盗版热线：(010)81055315

广告经营许可证：京东工商广登字 20170147 号

“大数据金融丛书”编委会

顾 问 陈 刚 刘文新
主 任 王玉祥 曹 彤
副 主 任 罗佳玲 刘文献 王作功 杨 东
编委会成员 简 毅 罗尧重 胡东婉 吴红军 陈宗权 王大鸣
李忠祥 朱志刚 刘建华 刘文献 梅 林 王恒壮
艾文华 姜 安 王宁桥 曹 峰 周 沙 武源文
栾明月 李梓正 杨 锐 张 冲

本书编委会

主 编 井底望天 武源文 赵国栋 刘文献

副主编 方 亮 顾善清 蒋晓军 梁 栋

撰 稿 张文彤 朱 立 侯毅瑾 蔡维佳 张新宇 许中阳

陈小虎 杨建新 李正鹏 张 玥 王磊磊 李忠阳

周 阳 石 乔 沈学峰

自序

大数据产业发展到今天，取得了令人振奋的成果，几乎所有领域都被大数据的热潮席卷。我们看到，随着互联网信息化技术的快速发展，以及云计算、物联网技术的逐步成熟，海量数据的生产、采集、存储和处理都不再是问题，数据的潜力得到了空前的释放。大数据时代并不是瞬间诞生的，而是信息社会发展到一定阶段必然会出现的结果。从长远来看，大数据会给整个人类社会带来了天翻地覆的变化。但是，这种变化需要通过多年的技术积累和社会实践才能达到，其发展过程也需要众多相关技术的同步发展和融合。

大数据价值的发挥在于多源数据的融合。目前的数据流通已经严重制约了社会整体大数据价值的发挥，数据的开放、共享、流通和隐私保护问题成为了大数据快速发展道路上最大的瓶颈。所有人都意识到，只有当不同的数据源开放共享，才能最终达成“社会化大数据”这个目标；否则，独立存在、互不共享的数据源只是形成了一个数据孤岛而已。但是，数据开放共享所面临的阻力可能远远超过人们的想象。究其原因，现在的信息化技术——数据库、云计算、数据中心等都是基于为中心化服务的思想而设立的，这必然导致数据高度集中，形成数据垄断。因此，数据垄断扼杀数据创新的问题也将长期存在，如何在数据所有权和数据共享之间找到合适的平衡点将是大数据

据生态能否健康发展的核心问题之一。

当中心化的云计算、大数据技术方兴未艾时，作为其对立面的去中心化的区块链技术已经横空出世，在虚拟货币领域风起云涌，这正好是中国人阴阳平衡思想的体现。在《区块链世界》一书中，我们专门阐述了不同于海外同行们所追求的“替代中心化的极端去中心化路线”的观点，坚定地认为区块链不是云计算、大数据等中心化技术的替代，而是有益的补充和平衡。正如古人所言，“阴在阳之内，不在阳之对”。中心化技术和去中心化技术是应该互补和相辅相成的，区块链技术与大数据技术必将完美融合，从而产生巨大的社会价值。

我们知道，区块链技术虽然随着比特币在 2008 年就诞生了，但是区块链这个名词却是 2015 年才出现的，最初大家是以大小写的 BITCOIN 分别表示比特币和其底层的技术。从 2016 年开始，区块链受到社会的广泛关注，由于其自身鲜明的特点和优势，一举成为了时下最新的热点，被称为与人工智能、虚拟现实并列的影响未来社会发展的三大技术方向之一，被认为是又一个革命性的互联网技术。区块链加密共享、分布式账本的技术特性对解决数据开放共享和流通提供了新的解决思路，不仅能促进数据的流通，破解大数据发展的困局，而且能通过构建价值互联网络逐步推动形成社会化的大数据互联互通。同时，区块链还能促进更平等和自由的数据流动，它所产生的基于共识的数据具有更致密的价值属性。因此，区块链很可能将成为一个提升人类社会信息精度的工具。

从这个意义来说，给区块链一个便于理解的定义并不容易。社会上关于区块链的描述：去中心、分布式记账、价值互联网的说法都对，但好像并不能完整地表明这是一个多么有革命性的技术。因此，我们首先尝试给区块链

一个符合其身份的定义。

区块链是一种由各种技术和通信协议组成的全新的互联网底层软件基础架构。其最早被用作比特币的底层技术架构，为比特币构建了一套完整的去中心化的货币发行、流通和支付结算网络。从技术角度来看，区块链是一种全新的、基于互联网的去中心、分布式数据库技术，是一个基于点对点的价值传输协议；从社会经济角度来看，区块链是在互联网上实现价值互联互通、构建共享经济的底层支撑技术。狭义地说，区块链是以时间为顺序相连的链状数据块结构，是一种分布式总账技术。广义地说，区块链是集加密算法、分布式数据存储、点对点传输、共识算法等多项技术于一体的新一代互联网基础技术，在互联网上可以支撑上层应用之间的互联互通和安全共享，被认为是互联网 3.0 的核心技术之一。

003

区块链技术的革命性被拿来与互联网的上一个革命性技术——TCP/IP 相并列。TCP/IP 用代码协议彻底打破了信息传递过程中物理空间、中心控制、时间跨度以及成本的限制，而区块链技术解决了 TCP/IP 所不能解决的信息传递的真实性问题，以及在互联网上进行价值存储和传输的问题。

区块链作为新一代互联网基础性技术，将会通过与大数据的结合以及互联网的应用逐渐传导至社会经济生活的各个环节，带来众多领域的模式创新，甚至会颠覆和重塑很多现有的商业模式、行业运行和治理体系，其目前的发展已经引起了世界范围内的广泛关注和各界的高度重视。作为一种迭代性的重大创新技术、一种全新的底层协议构建模式，目前大家普遍认为这项技术将推动互联网从信息互联网向价值互联网、秩序互联网的升级换代，彻底解决互联网环境下的信任问题，从而推动大数据和数字经济的快速发展。

区块链和大数据技术是伴随人类社会一步步共生进化而来。大数据的内

核仍然是统计分析，其背后的动力是人类对未来的精准预测；而区块链的底层逻辑是去中心化、自制、开放和透明，无论人们是否关注过这些深邃的底层逻辑，构筑怎样的世界观体系，现实对技术驱动力量的需求只有一个——效率！科技始终要服务于社会需要这个大原则，区块链和大数据相结合，将真正达到促进社会协同融合的目标。

2016年12月底，贵州省贵阳市政府发布了《贵阳区块链发展和应用》白皮书。这是贵阳市利用贵州大数据综合试验区政策优势，探索区块链技术对政务、民生、商务发展应用的宣言。白皮书中率先提出了“主权区块链”和“绳网结构”概念，在全球区块链发展的理论方面取得了创新。同时，贵阳市政府正在大力推动首批12个应用场景试点的落地。场景落地的过程是创新的过程，也是快速形成产业聚集、资本汇聚以及人才集聚的过程。在陈刚书记、刘文新市长的统一部署和推动下，贵阳市各届上下一心、众志成城，在政府数据开放、数据的安全隐私保护、区块链技术应用三个方面的发展上大步前行。

最后，请允许我向给予我们很大帮助的贵阳市王玉祥副市长表示衷心的感谢和深深的敬意！王玉祥副市长不仅于百忙之中抽出宝贵的时间一次次听取我们的汇报，不辞辛苦地参与到本书架构的反复修改中，还亲自撰文贡献了部分章节的内容和观点，直至出版前夜仍在提出修改意见。在此，我谨代表团队向王玉祥副市长表示最衷心的感谢！

武源文

2017年4月29日

目录

第1章 大数据发展的困局 //001

1.1 大数据的前世今生 //003

1.1.1 大数据的史前时代 //003

1.1.2 小数据时代 //006

1.1.3 大数据时代 //011

1.2 大数据的价值 //015

1.2.1 大数据的优点 //015

1.2.2 大数据的应用价值 //019

1.3 大数据实践的困境 //023

1.3.1 大数据实践之近忧 //024

1.3.2 大数据实践之远虑 // //030

1.3.3 无法回避的个人隐私问题 //035

第2章 迅速崛起的区块链技术 //039

2.1 区块链技术的由来 //041

- 2.1.1 什么是区块链 //042
- 2.1.2 数字密码货币的底层技术 //043
- 2.1.3 区块链家族的其他成员 //046
- 2.1.4 区块链技术的业务内涵 //049
- 2.2 区块链的商业价值 //052
 - 2.2.1 价值共享世界的 TCP/IP //052
 - 2.2.2 全社会智能资产化 //055
 - 2.2.3 对接机器经济 //057
 - 2.2.4 优化社会构成 //058

第3章 对技术的哲学思考 //063

- 3.1 大数据与不确定性 //065
 - 3.1.1 被牛顿刻画的世界观 //065
 - 3.1.2 急需打补丁的世界观 //067
 - 3.1.3 大数据是研究不确定性的技术 //072
- 3.2 区块链与信息基因 //074
 - 3.2.1 进化论 //074
 - 3.2.2 共生进化 //075
 - 3.2.3 技术的进化 //076
 - 3.2.4 技术与人共生进化 //078
 - 3.2.5 作为信息基因的区块链 //079
- 3.3 竞争力量的源泉 //082
 - 3.3.1 中心化的力量 //082
 - 3.3.2 去中心化的力量 //084

3.3.3 无关道德，只看效率 //086

3.3.4 竞争力的内在实质 //088

第4章 技术共生演进 //091

4.1 大数据的技术沿革 //093

4.1.1 大数据关键技术 //093

4.1.2 谷歌论文为大数据技术奠基 //094

4.1.3 Hadoop 生态 10 年见证大数据时代 //099

4.1.4 大数据技术发展趋势 //102

4.2 区块链的技术沿革 //103

4.2.1 区块链基础技术 //103

4.2.2 区块链更高层级的技术：资产互联 //106

4.3 大数据和区块链技术之同 //107

4.3.1 分布式数据库：HDFS vs. 区块 //107

4.3.2 分布式计算：MapReduce vs. 共识机制 //107

4.3.3 分布式和集中式技术的螺旋发展 //108

4.4 大数据和区块链技术之异 //109

4.4.1 两个技术处于不同的生命周期 //109

4.4.2 大数据和区块链的主要差异 //111

4.4.3 差异能否调和 //111

4.4.4 可相互借鉴之处 //112

第5章 互联互通，引领变革 //115

5.1 数据流通 //117

5.1.1 大数据经济来临 //117

5.1.2 数据流通的发展与预期差距较大 //118

5.1.3 区块链能带来哪些改变 //122

5.1.4 打造可信任的数据资产流通环境 //126

5.1.5 展望 //127

5.2 价值互联 //128

5.2.1 以比特币为代表的去中心化性 //128

5.2.2 匿名性 //130

5.2.3 支付便捷性 //130

5.2.4 不可篡改性 //132

5.2.5 可编程经济 //133

5.2.6 区块链重构大数据产业 //134

004

第6章 区块链技术在相关行业中的应用 //137

6.1 金融行业的变革 //138

6.2 智慧物联网 //142

6.2.1 机遇 //142

6.2.2 现状与痛点 //142

6.2.3 区块链能带来哪些改变 //143

6.2.4 具体应用方向 //144

6.2.5 现有尝试 //145

6.2.6 展望 //148

6.3 知识产权的管理 //148

6.3.1 机遇 //148

6.3.2	现状与痛点	//150
6.3.3	区块链能带来哪些改变	//151
6.3.4	具体应用方向	//153
6.3.5	现有尝试	//156
6.3.6	展望	//160
6.4	智慧医疗	//161
6.4.1	机遇	//161
6.4.2	现状与痛点	//164
6.4.3	区块链能带来哪些改变	//166
6.4.4	具体应用方向	//167
6.4.5	现有尝试	//169
6.4.6	展望	//172
6.5	智慧能源	//172
6.5.1	机遇	//172
6.5.2	现状与痛点	//173
6.5.3	区块链能带来哪些改变	//174
6.5.4	具体应用方向	//176
6.5.5	现有尝试	//176
6.6	供应链	//178
6.6.1	机遇	//178
6.6.2	现状与痛点	//179
6.6.3	区块链能带来哪些改变	//180
6.6.4	应用方向和尝试	//181
6.6.5	展望	//182
6.7	数字资产的管理与交易	//183

6.7.1 数字化实物资产的管理 //183

6.7.2 数据流通 //185

第7章 智能经济 //187

7.1 助力共享经济 //188

7.1.1 什么是共享经济 //188

7.1.2 共享经济与区块链 //190

7.1.3 共享经济的发展方向 //194

7.1.4 共享经济的多种尝试 //195

7.1.5 展望 //197

7.2 插上虚拟现实的翅膀 //197

7.2.1 又一个热点技术：虚拟现实 //197

7.2.2 结合了区块链和大数据的虚拟现实 //198

7.2.3 可视化沉浸让思维开始跳跃 //200

7.2.4 虚拟现实高效连接数据与人 //201

7.3 拥抱人工智能 //202

7.3.1 大数据孕育了人工智能 //202

7.3.2 让人工智能的价值真正流转起来 //203

第8章 “区块链+大数据”开启新时代 //207

8.1 “区块链+大数据”的发展路径 //208

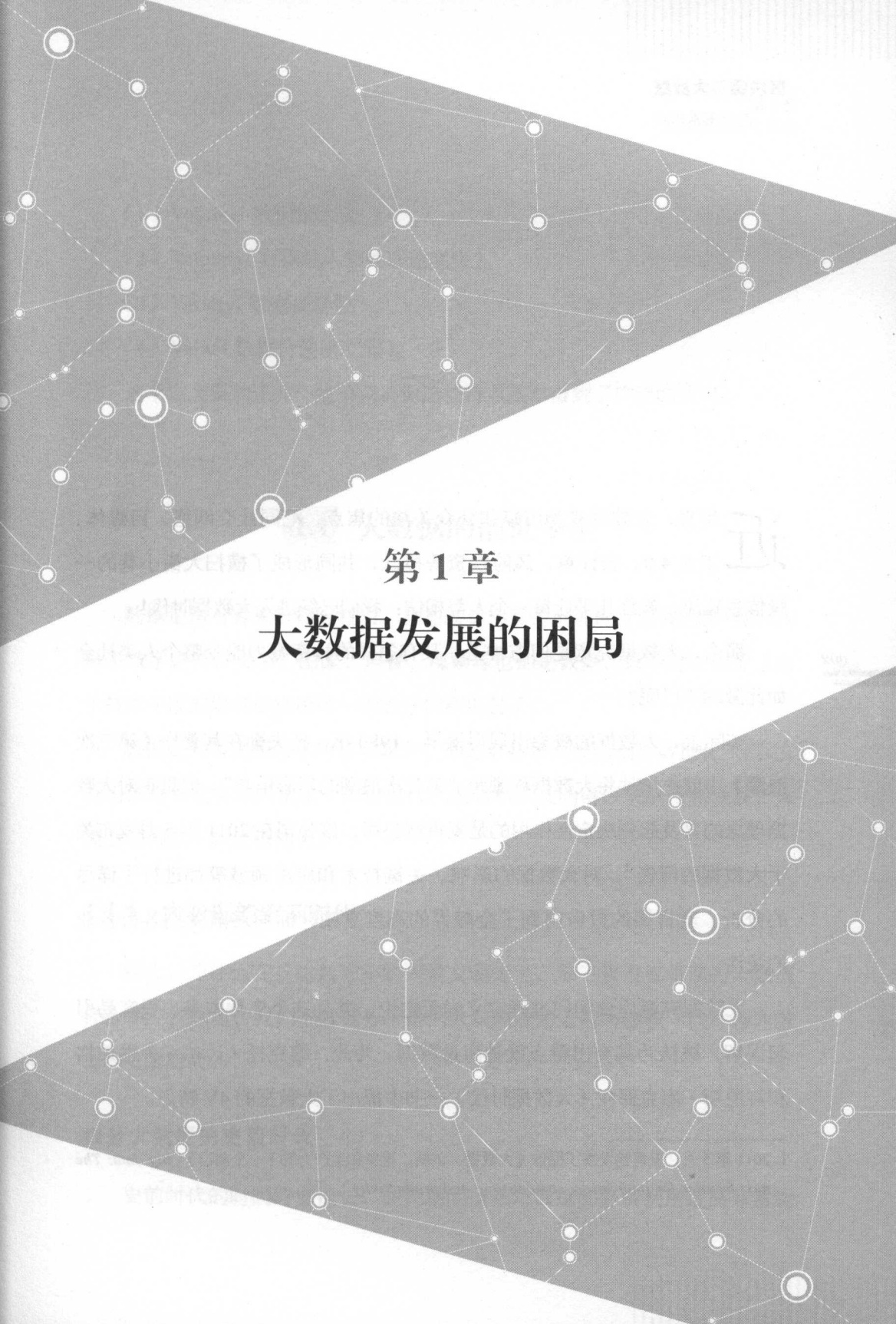
8.1.1 将区块链作为单纯的技术融入大数据采集和共享 //209

8.1.2 将区块链作为数据源接入大数据分析平台 //210

8.1.3 将数据作为一种资产在区块链网络中进行交易 //211

- 8.1.4 区块链作为万物互联的基础设施支持大数据全生命周期 //213
- 8.1.5 智能合约和大数据促进社会共治 //214
- 8.2 “区块链 + 大数据”实现新的社会融合 //216
 - 8.2.1 对旧体制的反思和对新方向的探索 //216
 - 8.2.2 新一轮的社会改良和变革 //219
 - 8.2.3 孕育新的经济发展方式 //220
 - 8.2.4 完善政府治理能力 //223
- 8.3 “区块链 + 大数据”共创新天地 //225
 - 8.3.1 政府层面推动是前提 //225
 - 8.3.2 企业是主体和核心 //228
 - 8.3.3 热门新职业：数据科学家 //231





第1章

大数据发展的困局

近年来，大数据成为传媒和公众关注的焦点，汇同社交网络、自媒体、工业 4.0、云计算、风险投资等概念，共同形成了横扫大街小巷的一股信息狂流，最终几乎让每一个人都相信：我们已经进入大数据时代！

那么，大数据究竟是何方神圣，又有怎样的神奇魔力能令整个人类社会如此激动不已呢？

实际上，大数据的概念出现得很早。1980 年，托夫勒在其著作《第三次浪潮》中就热情地将大数据称颂为“第三次浪潮的华彩乐章”。但真正对大数据理念的普及起到决定性作用的是麦肯锡公司，该公司在 2011 年 6 月发布关于大数据的报告¹，对大数据的影响、关键技术和应用领域等都进行了详尽的分析。麦肯锡的报告得到了金融界的高度重视，而后逐渐受到各行各业的关注。

大数据究竟应该如何准确定义尚无定论，并且这个名称本身也很容易引起误解，被认为其突出特点就是海量数据。为此，维克托·迈尔-舍恩伯格和肯尼斯·库克耶在《大数据时代》一书中提出了大数据的 4V 特点。

¹ 2011 年 5 月，麦肯锡发表了报告《大数据：创新、竞争和生产力的下一个前沿》（*Big data: The next frontier for innovation, competition, and productivity*）。

- (1) Volume(数据量大)。
- (2) Velocity(数据输入和处理速度快)。
- (3) Variety(数据多样性)。
- (4) Value(数据价值密度低)。

这个定义虽然还有一些争议,但这些特点基本得到了广泛的认可。

1.1 大数据的前世今生

数据是指对客观事物的性质、状态以及相互关系等进行记载的符号组合。除了数字之外,文字、图像、声音、视频等也都是数据,只不过在历史上由于技术手段的限制很难做进一步的分析利用罢了。

数据是人类文明传承的重要媒介。通过对各种人类活动的数字化(例如,利用语言和文字加以记录)和世代代的累积,文明得以被继承和发展,人类活动得以不断地精细化。

1.1.1 大数据的史前时代

那么,为什么在长达数千年的人类文明史上,却一直没有所谓的大数据什么事呢?这里,我们把数据分析得到大规模应用之前的漫长年代称为大数据的史前时代。

细分大数据的史前时代

史前时代的时间跨度很长,在此期间虽然对数据的分析利用进展非常缓

慢，但也逐渐出现了一些变化。史前时代可以大致分为以下几个阶段。

➤ 远古时代

远古时代是指文字系统尚未出现的时期，信息交流主要依靠口耳相传，如某种技能的师徒传承。虽然在后期逐步出现了结绳记事之类的简单记录方法，但总体而言，数据的准确记录都无法得到保证，就更无法考虑数据分析利用的问题了。

➤ 文字时代

文字是客观记录人类互相交流信息的一种方式，而不是一般认为的仅仅是记录语言的符号。文字的产生在人类文明史上起着至关重要的作用，它的出现意味着前人得到的知识有可能被准确无误地传承下去。目前世界上公认出现最早的文字是距今 5000 ~ 6000 年前两河流域的楔形文字。此后，世界各地的不同文明分别发展出了各自使用的文字，如古埃及象形文字、美洲玛雅文字、中国甲骨文、古印度印章文字、腓尼基字母、希腊字母、拉丁字母、斯拉夫字母等。文字的出现初步解决了数据信息的保存问题，也使数据信息的交流变得可行和可靠。但是，由于各个文字体系并不相同，跨文明间的交流仍然是非常困难的。例如，对数字的记录和计算方式，中国古代采用的是十进位制的算筹记数法，这在历史上非常先进，也使古代中国的数学取得了许多卓越的成就；古罗马的数字系统没有位值制，只有七个基本符号，如需记录稍大一点的数目就相当繁难；古美洲玛雅人倒是懂得位值制，但用的是 20 进位；古巴比伦人用的则是 60 进位。缺少统一且高效的数学语言，使在文字时代对数据信息做有效挖掘仍然是非常困难的。

➤ 阿拉伯数字时代

一般认为是公元 3 世纪，印度科学家巴格达发明了阿拉伯数字，并在大

约公元9世纪传入阿拉伯地区，随后传入欧洲，大约在公元13—14世纪传入中国。由于阿拉伯数字笔画简单、书写方便，加上使用十进位制便于运算，逐渐在各国流行起来，最终成为世界各国通用的数字。数学语言，特别是计算进制的统一，终于使各个文明之间能够对数据的分析利用进行有效的交流。此时，数据记录、存储、分析、交流的前提条件都已具备，数据的分析需求逐渐产生，小数据时代开始逐步向人类文明走来。

大数据史前时代的基本特征

■ 除非人工特意记录，否则不会自动产生和保留数据信息。这一时期并不存在对数据进行自动观察和记录的仪器，除非有人站在十字路口计数，否则就无法得知当日通过该路口的人数。而这个数据如果不被记录下来，也很快就会被遗忘。显然，数据在这一时期是异常匮乏的。

■ 严重缺乏对数据信息进行加工利用的技术和工具。数据分析高度依赖统计学，而统计学是数学发展到一定阶段才会产生的分支。在阿拉伯数字尚未普及的年代，罗马和希腊的数字用于计算加减都非常困难，计算乘除更是难如登天。

■ 严重缺乏数据信息交流的工具和手段。在文字出现之前，数据信息的交流只能靠口耳相传，如祭司师徒之间的知识传承或者游吟诗人的世代传唱。而印刷术出现之前，则只能靠手抄文字加以流传。直到印刷术出现之后，信息才有了大规模普及和交流的基础。但这种印刷层面的媒介尚远远不能满足知识本身的传播需求，更不要说加以二次分析利用了。

以上客观条件的限制,使有意识地采集数据并利用数据达成某种分析目标几乎不可能。传统文学作品中所称颂的日出而作、日落而息、男耕女织的生活,其本质就是完全靠天吃饭,没有足够的信息可以依靠,很难提高生产力水平。

信息蕴含着宝贵的价值,如果信息交流不畅,知识无法共享,就会导致各地的文明之间不能互相学习和借鉴。勾股定理就是个典型的案例,它被各个人类文明独立地重复发现,从而有了毕达哥拉斯定理、商高定理、百牛定理、驴桥定理和埃及三角形等多个名称。

文明总是在不断进步,人们逐步开始意识到信息的意义。即使是在大数据史前时代,也确实存在采集数据、分析数据的需求,这方面以作为社会管理者出现的政府的需求最强烈。据说古埃及就进行过人口普查,古罗马在拥有数十万人口的时候规定每5年普查一次。然而,普查所需要的能力远远超越了当时人类社会的上限,以至于随着国家人口越来越多而最终只能粗略计数。1086年,英国国王威廉一世对全国进行人口普查。结果到1087年他去世时,这项工作也还未完成。实际上直到19世纪为止(此时已经入小数据时代),即使这样不频繁的人口普查依然是很困难的。美国在1880年进行的人口普查耗时8年才完成数据汇总,因此他们获得的很多数据都是过时的。

1.1.2 小数据时代

虽然非常漫长,但史前时代终将过去;虽然非常困难,但数据的价值终将显现。到了近代,随着上述几大障碍的逐渐弱化,数据信息的价值逐步显现,人们开始针对特定研究问题有意识地进行数据信息的采集,并最终

进入了开始利用数据的一个新时代。

细分小数据时代

根据技术、工具和理念的进化过程，小数据时代也可以大致分为两个阶段。

➤ 古典小数据时代

这个时代的特征是人类开始形成完整的数据分析方法论并加以实践，但是缺乏强有力的工具（如计算器或者计算机）来协助完成整个流程。

实际上，很难界定史前时代和古典小数据时代的分界时间，或许最早的时间点可以被界定在开普勒进行数据分析这一代表性的事件前后。很有才干的天文学家第谷·布拉赫（1546—1601年，丹麦人）用持续20年的时间观察记录了750颗行星的运行数据，位置误差不超过 0.67° ，却没有对这些数据做深入分析。而能力相对一般的开普勒（1571—1630年，德国人）则花了16年（1596—1612年）对第谷的观测数据进行分析研究，得到了开普勒三定律。这个研究项目持续了将近40年才得到结果。

在古典小数据时代，上述几个数据应用障碍坚不可摧，人类几乎只能赤手空拳地应对所出现的数据分析利用需求。

➤ 计算机时代

这个时代的持续时间并不长。1951年，世界上第一台商品化批量生产的计算机UNIVAC-I投产，计算机从此从实验室走向社会，由单纯为军事服务走向为社会公众服务。这一年被认为是计算机时代的真正开始。

计算机的出现首先受益的是数据存储（虽然可能是用最原始的打孔纸片来记录数据）和共享，而20世纪60年代数据库系统的出现更是大大提升了计算机在这方面的能力。1961年，通用电气公司的查理斯·巴克曼（Charles

Bachman) 成功开发出的世界上第一个数据库管理系统——集成数据存储 (Integrated DataStore, IDS), 可算是这一领域的开山之作。其次, 随着软件体系的逐步完善, 计算机也开始提供越来越强大的数据分析和展示功能。1968 年, 全球最早的统计分析软件 SPSS 的问世, 意味着数据分析的计算机时代已经初露曙光。随着工业领域、商业领域开始大规模使用计算机控制和计算机数据采集, 越来越多的数据开始经由传感器-计算机这样一个体系被自动或半自动获取, 加速了数据采集领域的变化。

显然, 在计算机时代, 整个数据价值被发掘、利用的体系都越来越依赖于计算机, 但是由于方法论上仍然遵循着古典小数据时代的基本逻辑, 因此仍然属于小数据时代。然而, 各方面的不断进化都意味着传统的分析方法论已经过时, 一个新的时代即将到来。

分析方法论的更新

在小数据时代, 数据的价值逐渐被充分认识。随着科学的进步, 对数据进行加工利用的技术和工具被系统地发展起来, 最终形成了针对数据收集、整理、分析、利用的一整套方法论。

具体而言, 研究一个问题一般需要经过以下基本步骤。

➤ 确定研究目的

在整个数据分析流程开始之前, 研究者必须首先明确自己希望解决的问题究竟是什么。随后的整个流程都需要围绕达成这个目标而展开, 不相关的因素则被剔除。

➤ 研究设计

由于数据的采集需要成本, 因此研究设计的基本目的就是尽可能高效地

使采集到的数据信息能够集中反映所希望研究的问题。研究设计中会充分考虑如何尽量去掉不必要的样本量,有哪些需要控制的影响因素,并采用各种精巧的设计方案来对非研究因素的作用加以控制,如配伍、完全随机抽样、随机分组等。

➤ 数据采集

数据在研究设计完毕后开始采集。如果可能,则整个试验过程会在尽可能理想的情况下进行,从而在试验或数据获取过程中也会对无关因素的作用加以严格控制。例如,毒理学实验中可以对小白鼠的种系、周龄、生活环境、进食等作出非常严格的设定。当然,尽量严格的试验条件控制也意味着每个原始数据的成本都非常高昂。

➤ 数据准备

采集到的原始数据往往存在一些错漏,也不一定符合随后的统计分析需要。因此,首先要对数据进行清理,对变量进行转换,对异常值、缺失值等进行必要的处理。

➤ 数据分析

数据分析的目的就是基于研究设计,采用恰当的分析方法将样本中蕴含的有效信息提取出来,用于回答最初的研究假设。针对如何高效地发掘样本数据信息,统计学已经发展出了一整套方法体系。其中的很多概念或方法,如方差、百分比、回归分析、相关性分析等,都已经成为了普通人熟悉的词汇。

➤ 结果报告 / 应用

基于分析结果给出研究结论,并提出可供操作的业务改进建议,如对人群进行戒烟宣传、要求在食用盐中加入碘元素等。

面临的数据应用障碍

相比史前时代，小数据时代已经开始了对于数据信息的大规模利用。但是，史前时代的三个障碍中只有分析技术得到了解决，数据产生和数据流通上的障碍仍然存在。具体而言，在实际操作中会发现以下问题。

► 数据采集成本高昂

在小数据时代，由于绝大部分数据需要专门采集，并且整个流程都需要专业统计人员的参与，导致数据获取成本很高。因此，必须采用抽样的方式来控制样本量，以使用尽量少的样本得到对研究总体尽量好的代表性。但是，抽样方式的使用意味着或许对于整个总体而言，统计信息是充足的；而一旦细分到某个亚群体，相应的统计信息很可能就不够充分，导致无法得到有价值的结论。

► 数据共享和重复利用困难

高昂的数据采集成本会促使原始数据信息被严加保护，数据分享变得异常困难。不仅如此，由于每个研究的具体目的不同，样本选择的标准以及具体采集的信息也并不相同。因此，相应的数据几乎都只能为所对应的研究项目服务。除非在研究设计上专门做过考虑，否则将不同研究项目的数据互相打通加以重复利用几乎是不可能的。这也是为什么循证医学、meta 分析等数据二次分析的理念兴盛一时，但最后又归于沉寂的重要原因。

► 得出分析结论越来越难

在小数据时代，由于数据的来之不易，并且都是针对某个研究目的加以收集，因此研究假设的设定就要非常准确，这自然使研究结论会尽量往有因果确认关系的方向靠近。而随着研究的不断深化，比较简单明确的因果关系，

诸如吸烟会导致肺癌之类的事情大都已被研究完毕。找到明确的因果规律变得越来越困难,经常要反复经历“假设-求证-推翻-再假设-再求证”这样一个漫长的多次循环。再加上已有的数据信息很难被重复利用,这又进一步直接导致了研究成本的飞速上升。

► 分析结果的应用成本高昂

数据采集的障碍也使研究结论应用到个案上的成本很高。“我在广告上的投资有一半是无用的,但问题是我不知道是哪一半。”广告营销界广为流传的这句话反映的就是这种困惑。也许数据分析报告显示,某个保险产品的定位非常明确,目标客户就是“40岁以上,年收入30万元以上,喜欢外出旅游,在海外有投资账户,且为A型性格”的特定人群,但是由于数据的缺失(有的涉及个体隐私,有的必须专门采集,在营销上实际无法实现),使目标客户在营销时无法精确定位,只能采用低效而浪费的广告方式加以营销。

011

1.1.3 大数据时代

促成大数据时代的因素

► 数据的产生

近几十年来飞速发展的IT技术,特别是互联网技术,对催生大数据理念起到了至关重要的作用。长期以来,数据的采集成本始终居高不下。如今,这个障碍已随着多个自动数据来源的出现而渐渐消失。

数据的第一个来源是“电脑”。人类生产、生活的数字化让几乎每个使用电能的设备都有了至少一个核心处理器,此处称之为“电脑”。过去绝大部分系统运行的数据并不能被记录下来,而有了“电脑”之后,这些设备中内置

的处理器、传感器和控制器在运行时的情况都能以某种数据形式展现。最基本的运行情况会被自动以日志的方式加以记录，而更复杂一些的数据记录则可能是控制状态、异常事件报告等。

这种“电脑”带来的广泛数据化，使原本被舍弃的次要信息也能保存下来。例如，在电话需要人工转接的年代，话务员只会记录和费用有关的通话时长、电话号码等信息，而在现代的电信系统中，包括起止时间、通话内容在内的所有控制信息都能被自动记录。有了这样的数据基础，运营商对客户通话行为的把握就更准确，提供针对性的、个性化的服务也才有了可能。

数据的第二个来源是各种传感器。实际上，传统的摄像头也可以被看成是一种原始的传感器。传感器的特点是拥有一个唯一的识别 ID，同时它会根据外界提供的信号进行必要的信息处理，并发送返回信息。例如，摄像头会自动记录响应范围内的视频信息，并储存或回传至服务器；现在广泛应用的射频识别芯片（RFID）则会接收阅读器（扫码器）发出的无线电波，并反馈储存在芯片内部的信息给阅读器。传感器可以被用于大量的工作和生活场景：零售业结算、物流跟踪、仓储管理、可穿戴设备等，而现在被热议的物联网在数据采集端的实质就是各类传感器的大规模应用。

数据的第三个来源是将过去已经存在的以非数字化形式存储的信息数字化。例如，各种古籍文献，还有在过去不被认为属于“数据”的语音磁带、图片、视频录像带、历史档案、病历资料、设计图纸等。这些资料都以非数字的媒介形式存在，实际上很难加以分析利用。因此，对这些资料进行分析的第一个基础工作就是将其全部转为数字形式。这一部分数据在未来所占的比例可能会逐渐下降，但在现阶段仍然是数据的一个主要构成部分。

数据的第四个来源就是蓬勃发展的个性化互联网数据。在互联网时代之

前,前述三个数据来源实际上也是存在的,但其实施主体是企业。而在互联网时代,每个人都是数据的制造者,在微博上发送的文字和图片信息,在优酷土豆上传的视频及其讨论,在微信中的各种分享、聊天和互动,几乎全部都是由个体提供。这一部分数据占总数据量的比例正在迅速上升,在未来很有可能超过企业数据的总量。

还有很多传感器通过互联网实时采集来自个体的信息。例如,手机已经成为了个人信息中心,通过手机采集个体信息非常精确。现代的手机集成了GPS、声音、光照、运动、平衡等多个传感器,这些传感器所采集的信息(以及利用这些信息的各类App)在带给用户方便的同时,也精确记录着机主全天的详细行为信息,而这些信息的创造者就是机主本人。除了手机之外,日益发展的可穿戴智能设备也会提供越来越多维度的数据信息。

013

► 数据的传输

计算机软硬件的飞速发展解决了数据产生的问题,同时也提供了海量数据的存储方案。但数据本身其实是没有价值的,只有当数据流动起来,能够和其他数据或资源关联起来时,才会产生相应的价值。如果要让这些已产生的数据“活”起来,让其内部蕴含的信息得到发掘,则还需要解决如何传输分享这些数据的问题。

互联网技术的飞速发展最终冲破了这道壁垒。仅从网络传输速度来看,从20世纪90年代的拨号上网到现在的光纤入户,以及从有线网络到无线网络乃至3G、4G的进一步扩展,这些技术进步无一不在拆除着数据传输分享的技术壁垒。

数据传输问题的解决不仅使数据分享变得可行,也在数据采集和数据应用两个方面拓展了领域。例如,诸多无线传感器之所以能被用于数据采集,

都有赖于射频识别技术和后续近场通信技术的长足发展。这些技术的诞生也使很多业务场景能够实现，诸如仓储自动管理、电子支付、身份认证、票务管理等，而这些应用也反过来进一步促进了更多、更详细的数据得到采集。

大数据方法论

不同于小数据时代需要针对具体的研究目的特意去采集数据，大数据会在各种人类活动中从很多不同的来源自动采集进来。这些数据在生成时常常彼此独立，并未针对某种研究目的而刻意地去加工或者采集，完全处于自然采集的原生状态。

从统计抽样的角度讲，这些数据因其未遵循概率抽样原则，因此可能不具有很好的代表性。但是统计学也告诉我们，当样本量迅速增加时，样本对总体的代表性会自然增加（虽然相应的增幅会越来越小）。而大数据时代的一个显著特征就是海量数据，其数据量在理论上甚至可以逼近整个总体数据，如网民留下的网络行为数据。此时，人们会惊讶地发现，他们面对的不再是精巧设计之下的抽样样本，而是开始接近于研究总体的完整数据。

更重要的是，在小数据时代，样本只能对总体的一致特征具有代表性，而无法覆盖到各种亚群体的罕见事件（小概率事件）。随着大数据的出现，丰沛的数据量使研究者几乎可以将数据集任意细分到任何所关心的维度和深度，而仍然有充分的样本信息可被利用。这使得分析结果显然比小数据时代有了更好的代表性。

除了充沛的数据量之外，数据的多维度也使其代表性进一步增强。当不同来源的数据彼此孤立时，可能并无任何分析价值。但是，当它们能够通过某些关键点（如个体 ID）连接起来时，实际上就会从更多的侧面对目标个体

或群体进行立体画像，使基于这些维度对相应目标的推断更准确。

数据的代表性还体现在数据能够被多次重复使用。历史数据实际上就是对个体某种行为规律的记录，这些数据在采集时可能并无其他用途，但是累积到一定程度，特别是在和其他维度的数据相结合时，就可能揭示出更深入的用户信息。例如，移动电话运营商的数据库中会自动基于基站“握手”信息收集到用户的位置信息，这些数据当时只是用来传输电话信号，并无其他用途。但是，当它积累到一定程度时，显然就会揭示出该个体的日常行动规律。如果再进一步结合其通话业务量、通话行为特征等数据，就能够更加精确地定位出用户的消费特征，从而可以被一些发布个性化位置广告服务和促销活动的公司再次利用。事实上，滴滴打车等公司之所以要砸下几亿元来争抢用户出行的入口，其核心目的也就是为了获取这些有价值的用户行为数据。

1.2 大数据的价值

1.2.1 大数据的优点

如前文提到的，通常认为大数据有4V特点：Volume(数据量大)、Velocity(数据输入和处理速度快)、Variety(数据多样性)、Value(数据价值密度低)。除了价值密度低之外，其他三点都可以认为是大数据优于以前数据的方面。

数据量大是大数据最显而易见的特点，但绝不是大数据的全部。正如前一节所述，现代大数据的产生已经不再是专门设计和采样的数据，而是大规模

模自动采集的各种数据源、传感器和互联网数据。

不过，这样出来的数据不但是海量的和实时的，而且是极度差异化和鱼龙混杂的。毫无疑问，这种数据直接使用的价值比较低，且数据量极为庞大，因此需要经过筛选和处理才有可能发挥巨大的作用。

数据统计分析效率的突破

► 数据量大——对总体的完整描述

在小数据时代，由于成本和手段的限制，只能使用样本作为代表，而这必然意味着会有很多小概率事件无法覆盖。随着总体的逐渐细分，所对应的抽样样本越来越少，最终相应的研究结论在对应层级内的推断误差会越来越大，最终失去其应用价值。这被认为是传统数据分析方法的死穴。

而在大数据时代，由于数据采集入口的自动化，至少在某一领域获得无限接近研究总体的数据量是有可能的。例如，过去想采集全部居民的面部信息是几乎不可能完成的任务，但是在今天则可以做到。通过身份证管理系统、新生儿管理系统、学生证、士兵证、社保、驾驶执照等各种系统，就可以获得几乎覆盖全部居民的标准正面相。这样就不仅可以回答“中国人的典型相貌是怎样的”这样一个问题，还可以进一步地深入回答“上海出生、原籍四川的未婚‘90’后女性，其相貌特征是怎样的”这种高度个性化的问题。只要数据量足够完备，理论上对无限细分之后的层级也能够给出足够高精度的分析结果，这就使分析结论的代表性大大增强了。

► 数据多样性——全方位补齐缺失的核心信息

除了对总体做尽量精确完整的描述之外，大数据还可以全方位地补齐被刻意隐藏起来的、难以直接获取的关键信息，而这类关键信息往往都存在巨

大的商业价值或者其他价值（或许因此才会被刻意隐藏）。

能做到这一点是因为大数据的一大特点是多数据源，即数据采集范围不会再限定于指定的那些变量里，而是“漫无目的”地顺便收集各种各样的信息。因为变量之间多少都具有相关性，当某一个核心变量缺失时，只要集中采集了足够多与其相关的变量，就可以通过统计方法将该变量的数据以足够高的精确度估计出来，即使这些采集到的变量和核心变量只是弱相关性。

例如，每个人的收入信息显然属于个人隐私，当无法直接获取该数据时，在大数据时代完全可以通过许多间接变量对其加以推测估算，诸如其居住小区的档次、家庭居住面积、拥有的汽车品牌及价格、使用的手机型号以及手机使用行为、所就职的企业信息和相应职位、所担任的社会职务、拥有的各类会员卡、每年国内外旅行次数及时长等。通过对这些相关信息的不断细分和组合，实际上就可以对其收入范围做到非常精确的推测估算。在某种程度上，大数据的本质就是利用全面的信息来消除不确定性，这种特性是大数据完备性的体现。

017

更好地对未来进行预测

描述和补齐缺失值只是对现状进行呈现，而大数据的最终价值在于对未来进行预测。可以说，这方面的应用场景是充满想象力的。

例如，在公共安全方面，首先通过历史数据就可以预先得知在哪些节假日的哪些具体时间段，哪些公共场合容易出现人群过多聚集的现象，据此可以提前安排交通管制，调配警力资源进行管理。而当天更可以通过实时采集相应场所人流的手机移动信息，结合流数据技术进行实时分析以直接实现人流监控，并预测可能出现的安全隐患。这个监控过程中也会用到大

量的预测技术，如同步监控周边地铁、公交、私家车的流量情况，就可以提前一两个小时预知将来的人流密度，从而充分做到防患于未然。现在节假日的出行拥堵预测、旅游景点人流量预报等，就已经是在朝这个方向努力。而这几年“双11”快递送货速度明显越来越快，背后也都是基于历史大数据分析进行提前仓储配货、提前配置快递人力资源、做到物流最优化所带来的效果。

以大数据为基础进行的预测，小到体育比赛、电影票房、产品寿命，大到交通管理、流行病预测和社会经济发展趋势，不但对行业意义很大，更有巨大的社会效益。人力资源、物质资源、社会资源在大数据预测指导下的优化配置，极大地促进了生产力的发展和社会的进化。而经济效益最明显的方面，毫无疑问是金融领域，无数金融模型和分析都需要大量的大数据信息作为基础。

虽然现在大数据对未来的预测还远非完美，各种不匹配和“黑天鹅”事件还会层出不穷，但已经相当程度上展示了大数据预测在未来发展中的广阔空间。

分析结果可以迅速被投入使用

无论对现状的描述，还是对未来的预测，分析结果都需要能够用于指导实践，才会最终发挥其价值。但是，过去由于组织结构和技术能力的限制，将分析结果反馈至业务层面并产生作用效果所需要的流程非常漫长。

市场营销学有个经典的“啤酒和尿布”案例，即当零售商发现啤酒和尿布之间存在交叉销售行为时，需要重新组织货物的摆放方式，将这两种商品摆放在一起以获得更大的销量。由于它们分属于不同的商品组，将两个整组的商品都放在一起显然不太可能，基本上只能挑选出少量品牌以入口堆头的

形式放置，而这又会挤占有限的堆头资源。等讨论出最佳的商品陈列方式并调整完毕，至少也需要好几个工作日。但是在大数据时代，上述“应用-反馈-调整”流程完全有可能被大大缩短。最典型的是电子商务领域，商品浏览页面都会专门设置商品推荐栏，当发现某些商品存在关联销售时，只要用户浏览其中一种商品的页面，就会直接在相应位置显示关联商品的推荐，分析结果几乎在瞬间就可以得到应用。

这样不但可以大大缩短分析结果的应用流程，也可以明显提升结果应用的效率。由于可以补齐缺失的核心信息，并且数据可以精确定位到每个个体，从而可以实现对个体差异化的精确营销。也就是说，即使浏览同一种商品，A用户看到的关联推荐商品将会是基于A用户的属性特征给出的，和B用户看到的会完全不同。通过这种个性化的分析和推荐，用户的体验将会被大幅度提升，而浪费在广告上的那一半无用投资也将会被大大压缩。

019

1.2.2 大数据的应用价值

一种新的生产要素形式

大数据时代，数据不再仅仅作为生产过程的记载、劳动成果的度量，其本身也成为了一种生产要素，是一种经济资产。

数据代表了对现实的量化与抽象，所以大量的数据自然就给我们带来更多、更全面的观察世界的方式，而新的观察方式可以揭示以前没有发现的特性。由此可见，数据其实更像是一种矿产。在发现了铁的作用后，铁矿的价值才得以实现。同理，当数据的价值被不断发掘之后，自然会有更多制造、挖掘数据的冲动。数据在生产活动中的角色就升级为了生产要素的形式，成

为经济资产。

改良：对原有流程的优化

大数据的具体应用也不是一步到位的，在早期主要还是基于已有的生产、生活流程进行优化，并从中寻找新的价值点。以旅游出行为例，诸如携程、booking 等网站实际上并没有创造新的旅游方式，它们只是作为一个中间平台，将有旅游需求的客户和出行、住宿、旅行社等资源对接了起来，并收取高额的中介费用。但是，当数据积累到一定程度之后，这些中介网站就可以对客户进行细分，并有针对性地推送定制旅游产品。目前在携程网上由携程自营的旅游套餐所占比例已经越来越高，这一方面更好地满足了客户的需求，同时也提升了携程的利润。

类似的情形也出现在作为 C2C 中介平台的淘宝网，只是淘宝主要是面向卖家收费，买家用户在这方面的感受不深。

对于企业和产业来说，大数据也会产生深远的影响。例如，RightETA 软件综合天气、时间表、历史数据等信息，从根本上消除了飞机到达时间的预测误差，大大降低了机场方面因预测误差产生的成本。这是利用数据对现有运作模式改进带来的效益。因此，由于大量不可控因素的存在，任何没有达到数学最优化的过程都存在用大数据手段进行改进的可能。

突破：创造全新的生产生活方式

大数据资源的重复利用性和高维度特征注定使数据的拥有者迟早会发现数据中的隐含价值，并找到合适的方式将其变现，而这种变现就完全可能创造出全新的生产生活方式。

对于企业来说，当数据积累到一定程度，其对用户和市场也会有新的认识。“饿了么”从外卖订购平台起步，积累了大量的用户饮食信息之后，发现原来可以把菜品供应和原材料供应联系起来，于是就有了给餐厅提供食材的“有菜”平台，进而打通批发、物流、服务等一整套链条。这种做法其实就是依靠对已有客户的深度了解，利用大数据带来的高效率逐步侵占与现有服务相关的上下游产业的模式。可以想象，这种扩张不会仅局限于同一产业的上下游，横向扩展只是时间问题。例如，饮食习惯不健康的人群反而可能对健康产品有更低的抵抗力。对大数据的深度挖掘会揭示其内在的深层含义，进而提供全新的商业机会。

从另一个角度看，电商对传统商业的冲击最开始来自于较低的成本（减少门面等固定开销），但最重要的差异还是因为电商掌握了传统商户无法获得的用户数据，并能依赖数据迅速调整。这个思路甚至可以拓展到原本并不从用户获取数据的企业。例如，文章写作从纸质媒体时代的单向发送到今天点击量的追踪。实际上，如果能对读者阅读时的生理、心理进行测量（如心跳、眼睛焦点的移动、瞳孔放大等），那么未来的文章写作甚至都可能由数据直接产生。

“饿了么”的案例更多体现的还只是拓展企业自身的业务链，而基于大数据的互联网征信则是开创了全新的业务蓝海。在现代社会，个人信用是非常重要的基础信息，但传统的征信方式反馈速度慢、成本高，无法满足互联网金融业务的需求。而基于线上、线下各种维度数据源的互联网征信，将极大地丰富传统征信数据，又具有实时性的特征。基于母公司自身拥有的大量用户行为数据，蚂蚁金服旗下的“芝麻信用”以及腾讯旗下的“腾讯征信”明显拥有起跑线优势。而随着数据的逐渐积累，相应的模型不断迭代，其优势

很可能还会变得越来越明显。

颠覆：彻底重塑人类社会

既然大数据测量的参数涉及方方面面，那么这些参数也就可以被视为社会生活的脉搏，综合起来反映的是整个社会的即时运营状态。当对社会状态的把握进化到了对个体的实时了解时，社会模式创新就不可避免地会产生，包括且不限于国家治理模式、企业决策、组织和业务流程及个人生活方式等。

从国家治理规划的角度来说，原本的数据决策来自于对整体的粗放把握，社会数据化后对个体的反应会有较为精准的把握和预测，这对任何公共政策的制定和施行都是很好的辅助。同样，城市规划、医疗、教育等公共服务的提供也都可以做到动态调控，治安、消防、防疫等城市功能也可迈上一个新台阶，在实时和预防方面提高效率。

对于个人而言，大数据也将彻底改变其工作和生活方式。所谓工业 4.0，其本质就是自动化和信息化不断融合的过程，就是大数据持续发挥价值的过程，也是用软件重新定义世界的过程。一切都在基于数据被精确地控制当中，人类的大部分体力劳动和脑力劳动都将被机器人和人工智能所取代，汽车司机、售货员、检票员等工作将彻底消失，智能生产、智能服务、智能维护将贯穿于整个人类社会，不仅个体的生活将被重塑，传统的产业链也将被彻底打碎重组。例如，手表厂商届时可能会完全基于用户需求定制出每一个产品，而且免费赠送给用户使用。这个表每天贴着你的身体，采集身体的各项数据，这些数据对于保险业务就是个金库，手表厂商可以基于这些数据定制保险业务，那么手表厂商实际上同时也是一个保险公司。

对于人工智能来说，无论其决策机制如何，都必然要依赖数据的输入。下棋的人工智能需要应对的输入数据种类单一，而来自大数据的输入则可能以各种方式出现，必然会大幅提高人工智能的复杂度。如果说之前提到的大数据应用都是基于把相对通用的数据处理方式用在不同的个体身上，那么人工智能则是基于数据的个性而把对数据的处理进一步个性化，这么做最直接的结果就是更加精准的匹配与更高的满意度。随着时间的推移，掌握一切历史和当前数据的人工智能与其服务的用户之间甚至可能会产生一种类似教练与运动员之间的默契，真正进入数据驱动一切的状态。

简而言之，人类社会的一场深刻革命正在到来，现在还只是一个开端。

1.3 大数据实践的困境

近些年，大数据实践中最突出的问题就是基础数据的缺失。其原因一方面是没有统一的数据标准，因此很多数据无法被直接利用，需要转化或者清洗；另一方面是数据开放、共享机制不明确，无法有效地让已有的信息流通起来，供多方使用。

对于公共部门而言，独占数据能够保证自身的商业利益、信息安全，并且形成信息的不对称优势；对于私人部门来说，自身产生的数据往往涉及隐私，进而涉及个人隐私权的保护问题，但又希望得到个性化的定制商品和服务。那么，对自己信息开放到何种程度才是双赢的尺度？

所以，大数据需要面对的不仅仅是统一数据格式等技术问题，还包括数据质量、商业模式、社会伦理等一系列问题，更包括如何更好地利用社会资

源而不是陷入泡沫化的炒作。只有突破多方面的限制，大数据才会有长足的发展。

1.3.1 大数据实践之近忧

近年来，大数据的理念虽然很受追捧，但实际上还远未达到完全实用的程度。在市场咨询公司高德纳（Gartner）著名的技术成熟度曲线（见图 1-1）中，大数据概念也是 2011 年第一次上榜。

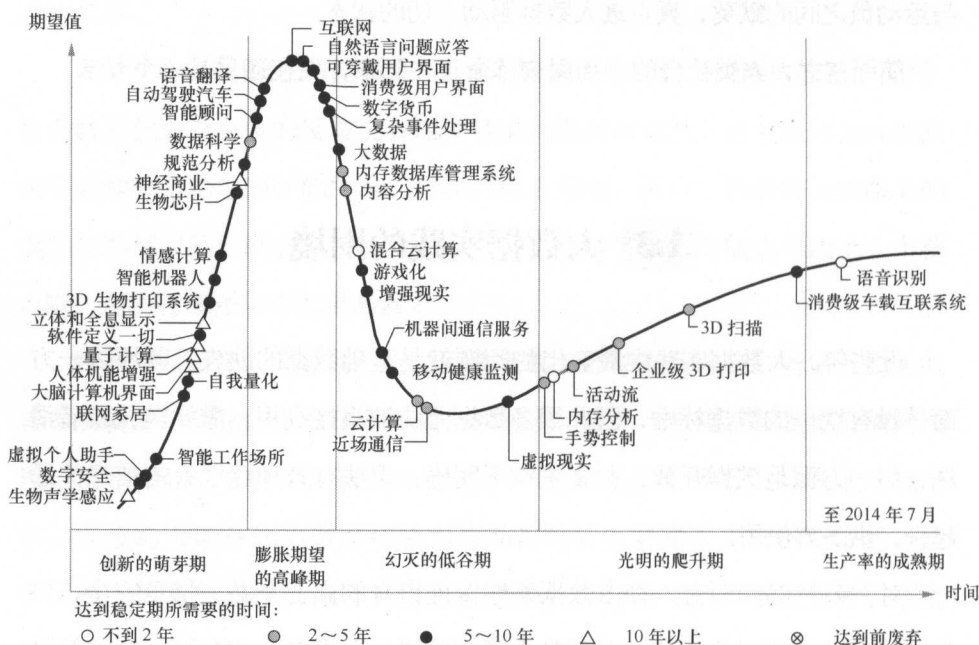


图 1-1 技术成熟度曲线

当时，大数据位于技术萌芽期的爬坡阶段，还被和其他概念整合在一起统称为“‘Big Data’ and Extreme Information Processing and Management”（“大数据”和极端信息处理和管理），随后就连续两年向泡沫期的高峰进

发。但在2014年,大数据在曲线中的位置就从泡沫期(Peak of Inflated Expectation)滑落,开始进入幻灭期(Trough of Disillusion),并在2015年被彻底剔除出该曲线。有趣的是,2011年高德纳在图上标明大数据还需要2~5年到达其稳定期,但是在2014年的曲线中,该时间则变成了5~10年。

那么,究竟是哪些因素在阻碍着大数据迅速实用化呢?

数据基础严重缺失

目前大数据发展最核心的一个问题,就是数据基础的严重缺失。丰富的数据源是大数据产业发展的前提,但是不同领域信息的数据化程度相差很大。对于制造、通信、IT等行业来讲,信息的数据化程度较高,虽然仍缺乏数据的共享和交换机制,但至少在各个企业内部已经开始进行一些大数据的探索和尝试。

但是在诸如医疗、教育等行业,进度则远远落后于大数据时代的需求。以医疗行业为例,这方面的大数据来源主要是电子健康档案,包括门诊治疗、住院治疗、健康体检、妇幼保健、慢性病管理等数据。从单个患者角度来看,自身的医疗数据包括储存在各医院的电子病历数据、公共卫生数据和居家自我监测的医学数据。但是,如果真正想将这些数据利用起来,就会遇到医院数据源不全、各信息源数据格式不统一、区域数据集中存储进展缓慢、数据安全和隐私保护等诸多问题。

中国医院协会信息化管理专业委员会的调研结果显示,三级医院信息化最高的部分是药库的管理,达到近79%,但客户管理信息化程度只有13%。而相当一部分医院连电子病历系统都没有,有的医院甚至还在使用手工缴费的方式,对这些医院的医疗数据进行整合是相当令人头疼的。数据标准化的

问题也不容乐观。2015—2016年医院采用统一信息编码体系的使用情况显示,使用度最高的ICD-10(国际疾病分类)也只有83.6%,影像传输的文件标准只有60%。标准不一也造成信息孤岛之间整合数据障碍重重,更使大多数数据成为沉睡的电子档案,开发利用无从谈起。

过度关注炒作带来的短期利益

大数据时代并不是瞬间诞生的,而是信息社会发展到一定阶段必然会出现的结果。从长远来看,大数据确实会给整个人类社会带来天翻地覆的改变,而这种改变需要通过多年的技术积累和社会实践来达到。但是,每当一个新概念或者新技术出现时,炒作总会比实干来得更加容易,获利也更迅速,大数据当然也不例外。

自2011年6月麦肯锡报告捧热大数据概念之后,几乎所有的商业领域都被大数据的狂热所诱惑,企业纷纷陷入了“如果不赶紧搭上大数据的车,就可能成为当代恐龙而最终灭绝”的恐慌之中。相应的商业和投资行为更多是源于炒作,而不是真正的需求所导致。当企业过了恐慌期之后,发现大笔的投资并未带来真正的回报,最终结果只会是动摇其对大数据的信心,并最终对大数据的实践带来伤害。

相关的IT软件、硬件企业和服务商纷纷加入炒作大军中争抢相应的订单。几乎所有和统计分析、数据采集、数据管理有关的企业,虽然所提供的服务内容并无任何变化,却在一夜之间全都成了大数据供应商,开始热情地向客户推销其经过重新包装之后的产品和服务。而在大数据的标签之下,真正实现的仍然只是很基础的数据仓库、动态报表展示、外部数据抓取等功能。

不甘落于人后的还有统计服务提供商,似乎所有的数据分析报告都必须

以大数据分析报告的面目才能示人，以致一个只是采集了几百例样本的传统数据分析项目在报告中也要冠以大数据报告的名称，并最终在媒体上以大数据的最新发现加以大肆宣扬。

当然，除了包装和炒作，也有企业在尝试基于大数据为客户提供的新价值。但问题在于他们对大数据的理解（或者说愿意做出的努力）本身就是非常有限的，以致能够实现的大多只是对社交网络或者指定网站的数据进行抓取、切词，最终以词云的形式加以展示，或者再进一步做出关键词提及频数变化曲线，就可以形成一个全新的大数据分析平台。但是，这些顶多只能算是数据采集和数据呈现而已，连大数据的初级阶段都算不上。更有甚者，这几年出现的很多基于大数据的聚合平台，也是采用类似的抓取操作，将网络上的相关信息整合成一个新的信息源，然后以 App 推送或者网站订阅的方式提供给用户，真正做到了“不是信息的生产者，只是信息的搬运工”。这类张冠李戴的所谓“大数据”服务，本质上也是一种搭车炒作，又能给用户带来多少真正的价值呢？能真正推动大数据行业发展吗？

027

盲目崇拜大数据干扰了正常的发展路径

这两年，炒作的盛行将大数据的魔力无限放大，大数据变得“一语胜万言”。只要是大数据分析得到的结果就无可质疑；即使是传统分析方法能得到的分析结果，也一定要用大数据方式再分析一遍，否则就被打上一个大大的问号。但大数据有其自身的边界，对大数据的盲目崇拜只会有害于这个新事物的发展。

大数据更多是辅助工具，而不是决定性工具。传统的数据分析方法虽然相对低效，但由于强调对因果关系的反复确认，使分析结论的准确性更高。

大数据由于更强调相关性而不是因果推断，使结果为假象的可能性明显上升。例如，某种大数据分析可以告诉我们从 2006 年到 2011 年，美国谋杀案比例与 IE 浏览器的市场份额极度相关，都呈急速下降的趋势。另外的大数据分析也可以告诉我们，儿童挨打的次数和儿童身高呈高度的正相关，但是两者之间显然并没有什么因果关系。

虽然从商业实践出发，当无法明确判断相关关系是否有实际意义时，基于这些相关分析的结论进行商业操作上的试探未尝不可，但真正的因果关系不仅需要严谨的逻辑论证，往往还需要在理论上找到相应的证据支持，甚至需要在相应学科的理论体系上做系统性的创新，这些显然都远远超出了大数据的能力范围。因此，由大数据获得的很多结果是需要返回来重新采用传统分析方法论加以进一步验证的，不可简单采信。

由于大数据本质上应当用来做出辅助性、探索性的结论，因此直接通过大数据来给出分析结论是有风险的。这也导致大数据分析实际上并不如看上去的那么有效。谷歌预测流感的案例曾经是大数据的经典，2009 年谷歌声称它可以通过分析与流感相关的搜索来预测流感爆发的趋势，这种准确性和即时性甚至超过了疾病控制和预防中心等官方机构。但是随后几年，谷歌的流感预测都没有得到好的结果，做得更多的是不准的预测。事后的分析认为，谷歌流感预测的失败很大程度上是因为它是基于谷歌用户的搜索行为进行分析，而用户的网络搜索行为本身就处于不断的动态变化之中。因此，谷歌搜索引擎也需要不断地更新，最终导致基于原先用户行为（以及相应数据）所建立的预测模型未必适用于下一个时期的用户行为及数据。

推而广之，大数据分析的内核仍然是统计分析，而统计分析的基本假设是历史数据中所蕴含的规律在未来仍然有效，因此可以基于历史数据的分析

结果来预测未来。但在用户行为变化速度很快的网络世界，这条准则很可能并不成立，因此所有依赖于网站用户数据的大数据分析都存在失效风险。

实际上，比谷歌流感预测失败更有说服力的可能是股票数据。股票市场的数据从股票上市以来就被完整、精确地加以记录，但是从来没有出现过能够精确预测未来个股走向的模型。绝大多数模型，包括近年来兴起的各种量化投资模型，用于拟合历史数据时其效果都是非常好的，但是投入实际运行之后往往迅速褪去光环，甚至给投资者带来惨重的损失。其实反过来思考一下就很容易理解，假如大数据真有那么大的魔力，那么大数据分析专家们直接去精确预测股市的未来走向就可以衣食无忧了，何必还到其他商业领域去挣辛苦钱呢？

对大数据的盲目崇拜可能导致一个严重的恶果，就是大数据导致大错误。由于大数据的结果应用经常强调时效性，而所使用的变量维度又很多，人力极难发现其中的虚假关联。而虚假关联的相互连接、相互影响有可能造成连锁效应，并在大数据的庞大量级之下扩大这些错误，最终导致严重的问题。最典型的案例仍然出现在证券市场。出于防范突发风险的需求，所有大的证券机构都设定了复杂的交易模型，以便市场出现突发情况时能够及时抛售证券来规避风险。这些交易模型尽可能地考虑到了各种突发情形，也基于历史数据做了很好的模拟，但最终导致的结果却是一次又一次的乌龙指事件。例如，在美国当地时间2010年5月6日下午2时47分左右，一名交易员在卖出股票时敲错了一个字母，将百万误打成10亿，结果引发所有交易模型的平仓行为，导致道琼斯指数突然出现千点的暴跌，创下了道琼斯指数历史上第二大单日波幅。显然，在大数据驱动的系统里，任何一个小的失误都可能被无限放大，造成严重后果，这在以前是不可想象的。

由于数据基础的缺失，相应技术尚不成熟以及对大数据理念的认识不深入，企业管理方式也无法迅速向适应大数据需求的方式转换等多种原因，造成了当前大数据应用中许多解决方案很难和实际业务深度结合。往往是有需求但无法获取相应的基础数据，有部分数据但很难和业务需求做到深度结合，有分析结果但很难在现有的业务流程中得到有效应用。

客观地讲，现在仍然处于大数据人才的培育期，大数据标准也尚处于构建期，需要有规模的大数据企业形成行业龙头，企业间建立统一的数据标准，打破数据孤岛，大量培养复合型人才，为大数据技术打下基础。只有这样，我国的大数据行业才能健康良性地发展，而这显然不是短期内能够做到的。

过度地关注大数据概念的炒作，或者满足于大数据泡沫化带来的短期效益，对大数据行业来讲是没有任何益处的。泡沫并不会解决大数据行业发展过程中的任何问题，只会带来一地鸡毛。

1.3.2 大数据实践之远虑

上一节所述的各种问题只是在短期内对大数据的发展造成阻碍，随着理念的深入、软硬件的逐步发展，长远来看这些问题都会逐步得到解决。但是，还有一些更深层次的问题将会在较长时期内一直存在，最明显的就是如何实现数据的真正共享。

数据孤岛之患

只有当不同的数据源开放共享，才能最终达成“社会化大数据”这个目标；否则，独立存在、互不共享的数据源只是形成了一个个的数据孤岛而已。但是，数据开放共享所面临的阻力可能远远超过人们的想象。

制约数据资源开放共享的因素有很多。对于政府部门而言，其本身就缺乏数据开放的动力。这并不能简单地归结为政府行政垄断、故意不作为或者理念落后等原因，政府在数据开放方面的保守态度是有其合理性的，因为政府所掌握的数据信息往往有一定的敏感性。例如，税务部门的个人纳税信息显然会涉及个人隐私；公安部门掌握的城市监控摄像头信息不仅会涉及个人隐私，更有可能涉及人身安全。至于政府手中的数据，即使是人口信息等基础数据，如果不加限制地共享也可能给国家安全带来严重威胁。这个问题在以“9·11”事件开场的21世纪就更是不能视而不见了。

除了政府部门之外，大数据时代的另一个主要数据持有者就是各大企业。但是，正因为数据中蕴含着宝贵的商业利益，甚至可能决定自身的生死存亡，因此各大企业也不会随便开放自身拥有的有价值的数据。以搜索引擎为例，众所周知谷歌的搜索效果要比其他的更好，但实际上他们的技术相差无几。真正造成效果差异的原因是谷歌的用户更多，积累的数据量也就更大，从而各种用户特殊的搜索需求都能在大数据中找到最佳的应对策略。而其他搜索引擎在这一点上就要差很多，进一步造成了其用户量始终上不去的恶性循环。基于大数据开放共享的理念，不同的搜索引擎应该彼此开放基础数据，从而都能为用户提供更好的搜索体验。但是很显然，至少对谷歌而言，这样的要求不会得到积极响应。

对于不存在企业间竞争的大数据场景，企业也会基于商业需求尽可能地将数据独占而不是开放。例如，对于各个航班的晚点信息，所有航空公司都不会提前披露，经常是当乘客登机结束之后才会在机场广播中告知航班将会晚点。从整体的航空运输角度，如果能够提前公开航班晚点信息并做出预警，显然会更有效地分流乘客，使航空客运的效率更高，也更节省乘客的时间。

但是，从航空公司的角度，公开这些信息将会造成有晚点风险的客源大量流失（如改乘高铁），造成公司自身的经济损失。因此，除非有强制性的行政命令要求，所有航空公司都会永远对晚点信息讳莫如深，甚至连历史数据都尽量隐藏起来，以最大化公司自身的利益。

破除数据资源开放共享的障碍，首先需要在数据保护和隐私保护方面构建完整的技术与制度体系，建立数据共享的基础；其次，各国应当有国家级大数据战略的宏观统筹，并在此基础上制定一系列的强制政策和法规，移除基于利益垄断需求而出现的各种障碍，以便最终能建立一个良性发展的数据共享生态系统。能否打破数据孤岛，成功形成一个开放的数据共享生态系统，可以说是未来大数据成败的关键。

低质量数据引发“劣币驱逐良币”

在小数据时代，统计分析圈里有一句格言：Garbage in, Garbage out。意思是说，如果采集到的原始数据是垃圾的话，那么分析结果也只能是垃圾。虽然在大数据时代，数据采集来源的极度丰富使不同维度的数据可以相互验证，因此对不精确甚至错误的数据有了越来越大的容忍度。但如果大部分数据源的质量都较差的话，最终仍然会使有效信息被淹没在无效的噪声背景之中。不幸的是，大数据时代由于数据生产、数据存储的成本飞速地降低，低质量数据的泛滥已经成为了一种常态。可以说，小数据时代的常态是缺乏信息，而大数据时代的常态则是信息极度丰富，但有效信息却被淹没在大量的噪声信息之中。

大数据时代噪声数据泛滥的原因，除了全方位的多数据源所采集信息的价值含量本身就非常低（但至少还是有价值可以发掘的）之外，还有大部分

是几乎完全没有任何价值含量的、纯粹扮演噪声角色的低质量数据。这些低质量数据主要有以下几种来源。

➤ 信息转发与复制

在门户网站刚刚兴起时，高度同质的文字会在短时间内泛滥于网络。而随着技术的进步，网络爬虫会 24 小时自动监控并抓取有用的信息，然后将其重新排列组合成全新的媒体内容并展示出来。这也就是近年来兴起的聚合类媒体所做的事情。在整个加工过程中，有效信息并未增加（当然，相应用户的点击、评论等操作仍然会增加有用的信息），但所需要分析的数据量和数据源却几十倍、上百倍地增加了。网络爬虫的历史至少和搜索引擎一样久远，也是谷歌等搜索引擎能够存在的技术支持。原本网络爬虫并不会造成太大的负面影响，但是在大数据时代，用户的每一次访问、每一次点击都会被作为数据资源记录下来供后续分析。此时，网络爬虫的负面影响就变得不可忽视。现在的很多互联网页面，每天都要被数以十万、百万计的网络爬虫搜索一次。而这些爬虫又会屏蔽真实 IP 来源，模拟人工浏览的各种行为，也使每个页面的点击量中究竟有多少来自真人、多少来自程序已经很难说得清楚了。

033

➤ 垃圾邮件与营销直投

这在全世界都是一个大产业，而其中则以垃圾邮件为代表。相信每一位读者都已经是垃圾邮件（SPAM）的受害者，它使大量的无效数据充斥人们的邮箱。充满讽刺意味的是，垃圾邮件之所以能够兴起，恰恰是因为大数据分析技术攻克了精确营销的难关所致，而大数据技术也进一步在帮助垃圾邮件（或者类似的营销直投信息）变得越来越难以识别。

➤ 网络水军

在购物网站兴起时，以刷帖为代表的网络水军就已经非常活跃。自从社

社交媒体出现之后，水军在互联网中的影响力更是大大超出了人们的想象。目前绝大多数网络热点背后都有水军炒作的影子。笔者曾经做过研究，发现在国内某社交网络中，新注册的用户至少 90% 可能都是“僵尸粉”。

在某种程度上，以上这些数据源都是数据垃圾的生产者。它们的生产速度显然会随着技术的进步而飞速上升，并大大降低真实有效的数据在网络中出现的频度，最终导致数据源整体质量的下降。

实际上，与垃圾信息识别有关的大数据分析技术会被攻防两方同时使用。这些技术能协助分析人员更准确地识别垃圾信息，但同时也可以使转发信息越来越像人工写成的作品。近几年，美联社、新华社等都已经开始用软件进行新闻的自动化写作。技术进步也可以用来使“僵尸粉”的行为越来越拟人化，其进化的速度远超想象。

这个问题对大数据的发展虽然不是致命的，但是如果不引起足够的重视，可能会严重阻碍其发展。

难以突破的创新瓶颈

大数据理念提倡的是数据开放和数据共享，却难以解决数据被越来越集中在少数几家巨型企业或机构这个现实问题。不但传统的大企业会垄断相应的行业数据，如通信数据在国内就完全控制在电信、联通、移动等少数几家有通信牌照的企业手中，而且即使最具活力的互联网企业，其数据源的垄断性也几乎是与生俱来的。例如，中国网民的网络行为数据大部分就掌控在 BAT 手中。社交网络数据可能稍微好一点，但也无非就是固定的那几家企业而已。

对于已形成垄断的企业而言，利用自身的垄断地位阻碍创新以使其垄断

地位更加巩固是很自然的事情。前文所述搜索引擎的竞争就是典型的案例。国内的情况也是如此，某大型互联网公司多年前就利用其已有的资源优势，采用跟随战略模仿潜在竞争对手的创新产品，并最终挤垮对手。某购物平台作为中国 C2C 市场的绝对霸主，实质上拥有了对相关数据制定规则和随意解释的特权。

从长远来看，由于数据所有权高度集中的问题无法简单解决，因此数据垄断扼杀创新这个问题也将长期存在。那么，如何在所有权和数据共享之间找到合适的平衡点，将是大数据生态能否健康发展的核心问题之一。

1.3.3 无法回避的个人隐私问题

每个人都正在成为数据洪流中的一部分，持续地使用、生产和分享着数据。共享已经成为新的数据道德特征之一。在大数据时代，数据素养开始成为公民的基本素养：对数据敏感，有一定的数据收集、分析和处理能力，有利用数据进行决策的能力，对数据具有批判性思维。

身份技术是大数据应用的核心技术之一。1993 年的《纽约客》发表了著名的漫画，主题是“在互联网上，没人知道你是坐在电脑前的一条狗”。时至今日，这个题目恐怕要改成“在互联网上，所有人都知道你是一条狗”。早期互联网的虚拟身份与现实身份之间有巨大的隔离，随着互联网深入渗透到生活的方方面面，虚拟身份与现实身份之间再也无法彻底地分割。各种账号系统、电脑中各种应用所放置的 Cookie、设备指纹技术、生物隐性行为识别技术（如 Biocatch）、图像识别技术等，都有助于用来识别出每个具体的人。尤其是当我们把社交关系搬到网上时，现实身份与虚拟身份实现了空前的统一。而广泛使用的社会化登录填平了跨屏的身份鸿沟，在网上留下了身份更

统一的“足迹”。于是，我们的身份背后出现了长长的“数字背影”。

场景分析是大数据应用的重要工具。场景分析就是研究个体的行为特征、心理特点和行动预测。场景化的产品应用就是基于不同的假设来试图了解和理解每个用户，从而设计出个性化的服务规则。基于场景化的大数据分析，系统可以深入洞察个体行为，让具体的分析具有实际的商业价值。因此，大数据通过场景分析实现了产品化，这些产品化的大数据可以对大数据进行高效的消费。于是，每个“数字背影”就变成了可以无限复制、具有高度流动性的、等待进入各类数字工厂进行加工的原材料。

之前的所有技术革新从未像大数据这样与个人隐私密切相关。《黑箱社会：掌控信息和金钱的数据法则》的作者弗兰克·帕斯奎尔认为，我们每天都要面对这样的问题：个人信息越来越多地被别人掌握，而每个人本身既不能阻止这种情况的发生，也不知道会产生怎样的后果。一方面，各种终端、传感器和记录设备无处不在地记录着我们的虚拟世界和现实生活中的应用数据和生活轨迹，通过大数据技术可以深刻地洞察我们的一切，甚至可以预测我们的行为；另一方面，作为这些数据归属权的主人，却不知道哪些数据如何被收集记录了，这些数据未来会出于什么目的、流向哪里、会被谁如何使用，这个过程对我们个人来讲就是一个神秘的黑箱。

共享数据、先予后取正在成为越来越多应用的必要条件。如果不允许地图导航应用使用所处的位置，你将无法获得准确的导航服务；如果不告知家庭住址和收货人信息，你将无法进行网上购物；如果不告知自己的个人兴趣信息，即使拥有社交媒体账号，系统将无法给你展示相关的各种信息流；如果不允许可穿戴设备上传量化数据，设备本身就等同于废物。

大数据技术必将会和个人隐私产生激烈的冲突，并将对社会产生深远的

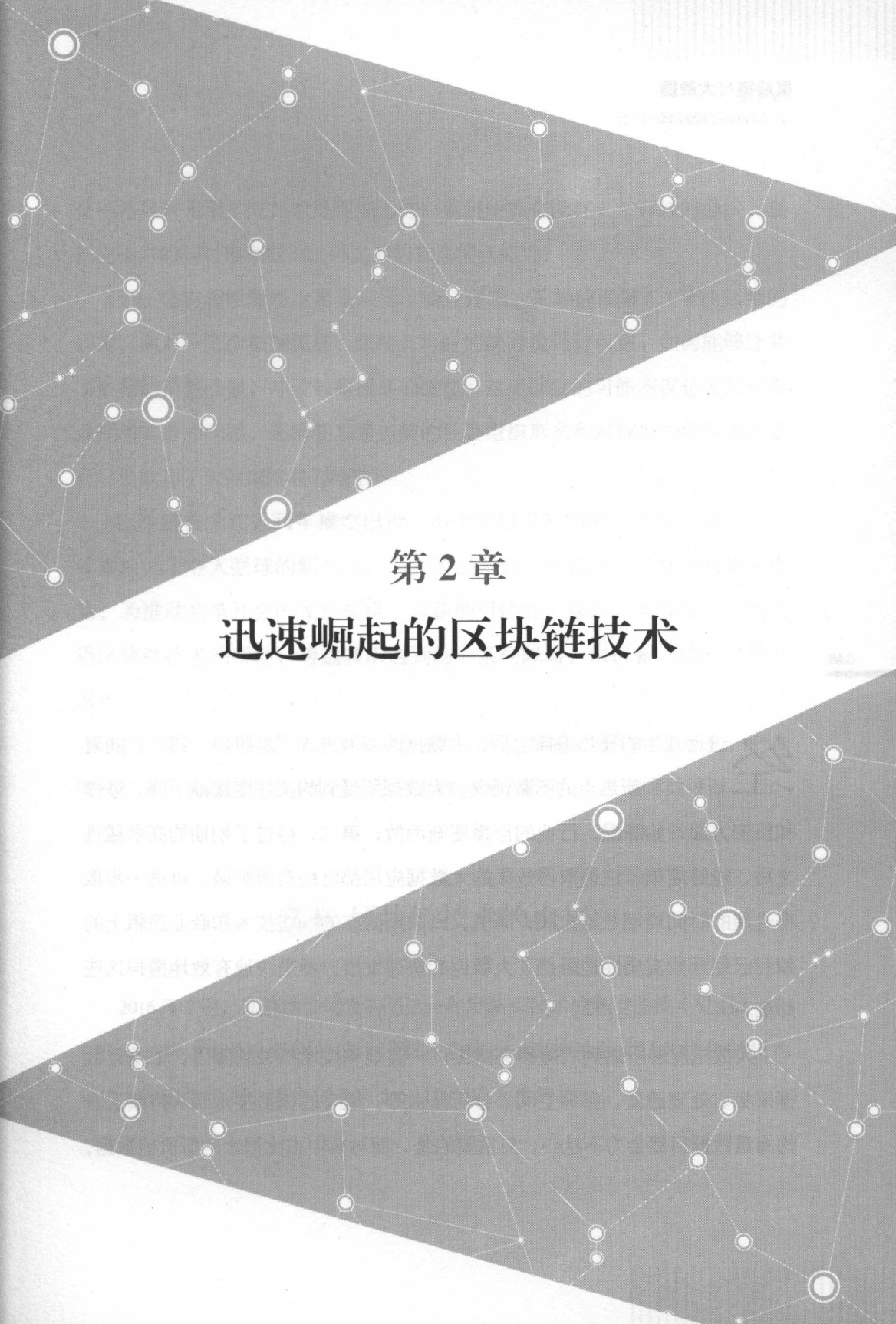
影响，如更广泛的身份攻击和盗用、不准确的数据与模型、不公平地使用敏感信息、算法影响个人行为、政府对公民控制能力的大幅增加等。早有预言说我们在网络时代将毫无隐私可言，但是谁也没有想到，我们的个人隐私在大数据还只是初露锋芒之时，就已经在广度和深度两个维度遭受到巨大的冲击以至于荡然无存。而且，大部分人却对此缺乏必要的认知。

大数据的发展必须解决个人隐私保护的问题。一方面，个人隐私不能够被无限制地滥用，每个人需要对自己的隐私使用有知情权，也有拒绝的权利。这种权利需要有可靠的手段，而不能仅仅建立在第三方信用背书和道德自律上（大量的电商和公共数据被盗事件早就证明了第三方的承诺并不可靠）。另一方面，包含部分个人隐私信息的数据又需要找到安全、可靠、高效的途径和方法来做共享，只有这样，社会化大数据才能真正发展起来。

037

那么，什么样的技术、什么样的商业模式和社会运转模式才能够支持个人隐私的合理使用呢？大数据在实践中又如何才能解决上述问题，找到合适的发展路径呢？





第2章

迅速崛起的区块链技术

经过近几年的狂热追捧之后，大数据的发展进入了新阶段：第一，随着新科技和新热点的不断涌现，大数据所受到的关注度逐渐下降，炒作和投资方面开始降温，行业的浮躁逐渐消散；第二，经过了初期的高歌猛进之后，能够简单、快速取得效果的大数据应用都已经有所突破，再进一步取得进展的时间周期开始拉长；第三，大数据面临的一些技术和商业逻辑上的掣肘已经开始实质性地阻挡了大数据的快速发展，亟需快速有效地搬掉这些绊脚石。

大数据发展所遇到的障碍有两类。一类是和技术相关的障碍，如基础数据采集、处理速度、存储空间、分析技术等，原有的相关技术面对持续上升的海量数据自然会力不从心。更重要的是，面对其中占比很大的低价值数据，

是不是只有无限扩充技术处理能力这一条出路？有没有更加有效的办法，在扩充能力的同时提高数据处理之前的数据筛选能力？

另一类实质性障碍主要是来自于商业逻辑。正如前面第1.3节所谈到的那样，面对一个个数据孤岛，纵使有再好的能力也无计可施。如何能够让共享数据的价值凸显，冲破数据孤岛的壁垒？这里面临的的可能不仅是大数据商业逻辑本身的问题，还需要思考当前的社会组织形态和对数据的使用模式是否已经成为了大数据发展的桎梏？

区块链技术在近两年横空出世，由于其自身鲜明的特点和优势，不但一举成为了夺人眼球的新热点，而且为很多行业难题提供了全新的解决思路，为推动共享社会的发展提供了全新的可能性。那么，大数据能不能获得区块链技术的帮助，跨越其所遇到的一系列障碍，取得新一轮的长足发展呢？

041

在回答这些问题之前，本章将会首先介绍区块链技术到底是什么，它有哪些优势，又如何与产业发展相结合发挥巨大的威力。

2.1 区块链技术的由来

2016年7月，高德纳公司发布了这一年度新兴技术成熟度曲线（见图2-1）。和往年一样，有几个新兴的技术出现在曲线上。高德纳公司认为区块链技术正处于期望膨胀期，将会在未来5~10年成为主流技术。



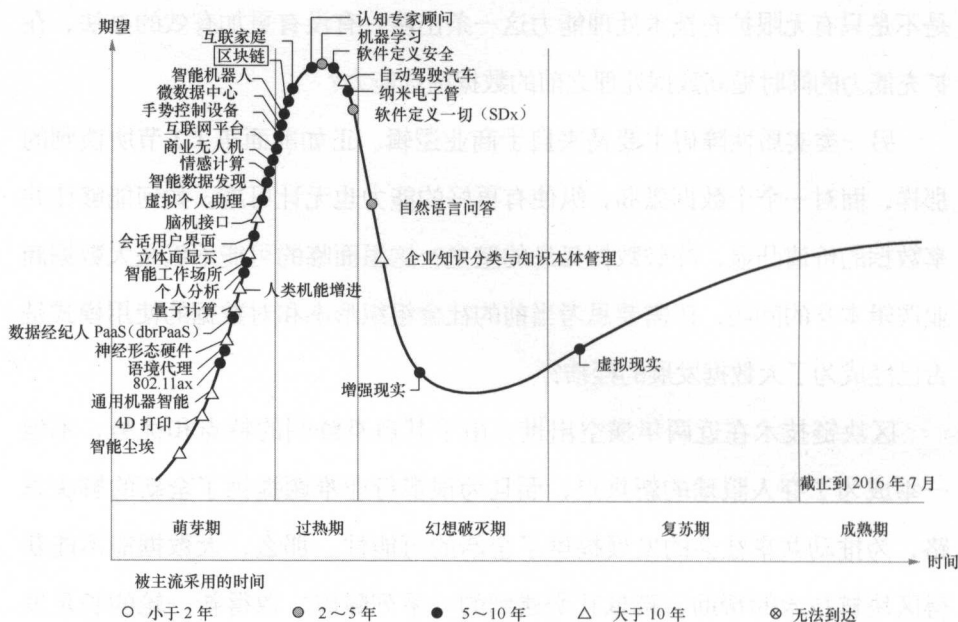


图 2-1 2016 年度新兴技术成熟度曲线

从技术角度看，区块链是“横空出世”的新星。但从价值逻辑层面而言，区块链蕴含的恰是“复古”的思想。古代人们比邻而居，生产力水平不高，尚未发展出银行等金融机构和社会化的大生产方式，日常活动更多地基于互信，如以物易物、就地生产和就地消费，这和彼此相连的区块链颇有相似之处。

随着生产力水平的不断提升，生产生活活动开始复杂化，一般等价物和金融机构应运而生，政府这样的社会管理系统慢慢成熟，原本扁平化的人类社会生活模式变得中心化。当然，在相当长的时间里，中心化带来了高效率，使人们具备了和陌生人发生多种经济活动的的能力，但是人和人之间的互信却越来越少。

现在出现的区块链，与其说是建立了全新的互信方式，还不如说其实是帮助人们回归了曾有的互信。

2.1.1 什么是区块链

作为一个新兴的信息技术，区块链使用时间戳和数字密码技术，把交易记录记载在按时间序列组成的数据区块中，并使用共识机制把数据存储到分布式数据库内，从而生成了永久保存、不可逆向篡改的唯一数据记录，达到不依靠任何中心机构而实现可信交易的目的。

区块链的数据记录是公开透明的，一个比较形象的例子就是在没有互联网和电话的年代，人们的远程信息交流主要是通过邮局寄信。区块链账本数据库和传统数据库的差别就好像那个年代明信片和平信的差别：明信片的内容可能会被邮局或周围很多人看到，收信人和发信人想否认也否认不了；而平信就不一样了，写的内容是什么，外人并不知道。

同时，区块链的数据记录是不可篡改、永久不变的。这个就好像多联复写的发票或收据，修改或者销毁单张单据无法改变其他单据的数据记录。

2015年5月27日下午17时左右，全国多地网友反映支付宝无法登录。下午18时许，支付宝通过微博发布消息，承认支付宝使用出现故障，原因是杭州市萧山区某地光缆被挖断。这个事件非常典型地反映了中心化数据库存在的巨大风险，无论是数据库本身遭致的损坏（机房出现问题、数据库数据被损坏或者篡改），还是中心数据库和外部终端的通信出现问题，都会导致中心化数据库不能发挥作用。而区块链的分布式存储就是把全部记录分布式保存在整个网络的多个记账节点上，单个节点的损坏或灭失并不会对其他节点造成影响，单个节点的数据错误或篡改更不可能对整体数据产生什么破坏性的影响。存储在不同节点的数据信息都受到密码学技术的严格保护，即使获得了相关信息，没有合法授权也无法偷窥到数据信息的真实内容。

区块链通过共识机制创造性地解决了全网各个记账节点的信息同步问题，可以有效地摆脱某些问题节点的影响，完成正确的记账更新。以比特币为代表的 POW(工作量证明) 共识机制还能通过提供有效回馈的方式，激励网络节点参与记账，吸引全世界的算力来为比特币当账房先生，从而有效维护了整个比特币区块链网络的运作和发展。

2.1.2 数字密码货币的底层技术

区块链得以为人所知，是因为它是作为第一个成功的数字密码货币——比特币的底层技术。

说起数字密码货币，比特币其实并不是第一个吃螃蟹的。比特币的许多核心概念都是在一个叫作密码朋克 (Cypherpunks) 的网络社区里发展起来的，这是一个聚集许多密码天才的松散联盟。密码朋克本身就是数字货币最早的传播者，在其电子邮件组中常见到关于数字货币的讨论，并有一些想法付诸实践。在比特币之前就有很多数字密码货币的失败尝试，如亚当·贝克的 Hashcash、戴伟的 B-Money、大卫·乔姆的 DigiCash 等。

中本聪在发给一个安全研究人员的信中诠释了比特币将如何避免重蹈这些前辈的覆辙：“你们知道，我想很多人对 20 世纪 90 年代的事情很感兴趣，但是经过 10 多年基于第三方可信任系统的失败 (DigiCash 等)，他们认为这是一个注定要失败的努力方向。我希望 (比特币这次) 他们可以做出不同的判断，因为这是我第一次知道我们在尝试一个不基于第三方信任的系统。”

2008 年 10 月 31 日，中本聪在一个隐秘的密码学讨论组上发表了比特币系统的白皮书《一种点对点的电子现金系统》，完整地阐述了比特币的原理以

及他对数字密码货币的看法。随后在 2009 年 1 月 3 日，中本聪“挖出”了比特币的创世区块，正式标志着比特币的诞生。

作为一种成功的数字密码货币，比特币的总量固定，而且有一定的发行规律，不会产生超发并引起通货膨胀，甚至还会有适度的通货紧缩。它的发行完全是公开、透明、可验证的，同时能解决数字密码货币实践过程中遇到的“双花”问题¹。

比特币的运行需要“矿工”进行类黄金挖矿（特殊的工作量证明机制，Proof Of Work），“矿工”们参与比特币系统的发行以及手续费的收取，每一个发行的比特币中都凝聚了“矿工”的算力。虽然“挖矿”被认为是浪费能源的（目前全球比特币系统的用电量每年超过 50 亿度，大致相当于现在一个中等城市的用电量），但是“挖矿”模式是迄今为止维持时间最长、运行状态良好的验证方式。比特币面对全球无数黑客的反复攻击，历时八年没有出问题，证明了数字密码货币是可信的。

也正是因为以上特点，比特币诞生以后迅速发展成越来越被人们所认知和接受的数字密码货币。行业内初创公司的融资甚至不再接受传统的法币，转而专门接受比特币的投资，证明比特币的价值获得了业内公认。传统的信用卡公司 VISA 也专门发行基于比特币的借记卡和信用卡，支持比特币的 ATM 机也日益增多，以至于财经类的分析中，专家们除了要分析各国汇率以及大宗商品的走势之外，还要分析比特币的币值走向。2017 年伊始，比特币的币值再次大涨大跌，又一次牵动了投资者和关注者的神经。特别是其天生

¹ “双花”问题（Double Spending Problem），也就是双重支付问题。比特币的解决方案是使用点对点的网络来解决“双花”。简单地说，比特币网络就像是一个分布式的时间戳服务器，为币的转移打上时间戳，这正好符合信息易于传播、难以消灭的本性。

具有的点对点转账和突破法币资本管制的特点，被很多人直接用于跨国资金流动。

区块链伴随着比特币的发展而成长。2015 年，区块链终于脱离了对比特币的依附而被公认为是全新的独立技术，可以在更加广泛的应用领域自由驰骋。区块链技术也天然具有比特币的几个特点：公开、透明、可验证，可以被广泛应用于社会经济、生活的不同场景，未来将成为推动商业模式变革以及社会发展的独特力量。

区块链最初的学术定义为由包含交易信息的区块从后向前有序链接起来的数据结构。例如，比特币的区块链示意图如图 2-2 所示。

区块链存储在由多个节点组成的分布式网络系统上，每个完整的节点都存有整个区块链的副本，而每一个计算机节点之间通过网络进行交易信息的共享。同时，区块链也是一个交易数据库，其中存储的是系统中由所有节点共享的信息，称为分布式加密总账本。通过这个总账本，区块链实现了不需要任何一个中央权力机构或受信任的第三方来协调互动、验证交易或监管行为的特征。

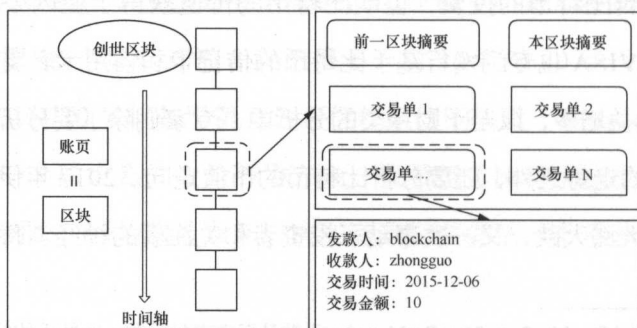


图 2-2 比特币的区块链示意图

一个区块链上的完整副本包含了每一个曾经执行的交易，使历史上的任何信息都可以被任何一个参加的节点所访问。

2.1.3 区块链家族的其他成员

区块链家族最著名的成员就是比特币，这在上一节已经做了介绍。比特币完整地运行了八年，从最初密码币圈子的“小玩意”走到了现在的声名显赫，从 50 个比特币换取一个匹萨到近期一个比特币价值 1300 美元（已经“挖出”的比特币总价值超过 100 亿美元）。比特币最大的意义就是面对全世界无数“矿工”的昼夜挖矿、面对无数黑客的反复进攻、面对各种流向的资金交易而屹立不倒，用事实证明了区块链技术是可行的。

跟随着比特币，大量区块链实践案例相继涌现。由于比特币并发处理能力不强，也无法快速确认交易，更没有智能合约这种大杀器，这些有后发优势的新成员在比特币区块链的基础上做出了各种各样的改进和优化，以适应广泛的社会应用需求。

比特币侧链

比特币团队为了商业化开发，提出了侧链解决方案。简单地说，就是在继续保持基于比特币的信用证明的同时，支持完成一些更复杂的应用操作。

侧链的原理是通过“双向锚定”机制实现主链货币（比特币）价值向侧链体系的转移，从而在侧链上使用这部分从主链转移过来的主链货币的价值。至于以这部分主链货币价值背书而产生、发行的侧链货币的名称，则可以按需自由命名。

闪电网络

为了提高比特币的交易效率，解决比特币的交易规模、实时支付、小额支付等关键应用问题，闪电网络（Lightning Network）就是其中最著名的一个解决方案。

闪电网络的基本思想就是交易建立在一个新的支付路径，支付路径中的所有交易只有最后一笔需要真实进入比特币区块链。其原理就是在森林里一棵树倒了，发出很大的声响，但是外部世界对此并不感兴趣。所以，很多高频小额交易对于比特币区块链并无意义，无需全部记录。

以太坊：开放、匿名的智能合约计算平台

以太坊（Ethereum）是一种新的去中心化区块链协议。2013年11月，当时18岁的俄罗斯裔加拿大少年维塔里克·比特林（Vitalik Buterin）创建了初始的以太坊概念和基本代码，他的目标就是将区块链技术所具有的去中心化、开放和安全这三大特点引入几乎所有能被计算的领域。

以太坊之所以能超越以往这些项目的局限性，是因为其核心理念很简单，就是一条内置图灵完备编程语言的区块链，允许在上面创建任何应用。以太坊要实现的是一个内置了编程语言的区块链协议：由于支持编程语言，因此理论上任何区块链应用都可以用这门语言进行定义，进而作为一种应用运行于以太坊的区块链协议之上，而不是像以往那些项目各自为政，分别定义自己的区块链协议。以太坊可以使区块链应用开发者高效、快速地开发顶层应用变为可能。

超级账本：建立联盟链、私有链的开源技术标准

超级账本（Hyperledger）是由非营利组织 Linux 基金会发起成立的、致力于企业级区块链开发及应用的开源项目。该项目的愿景是借助项目成员和开源社区的合力，制定一个开放、跨行业、跨国界的区块链技术开源标准，打造可以跨行业的区块链解决方案。

超级账本团队认为，支付系统在高度集权和完全去中心化之间应该有个平衡：权力既不是集中在某一个机构，也不是完全地分布式，而是合理地分割成若干部分。不同于比特币区块链的工作量证明机制让账本需要 6 次确认，每次 10 分钟，确认时间总计需要将近 1 个小时，超级账本则是采用类似 Ripple 的“共识”机制，交易确认过程可在几秒钟之内完成，达成共识则是通过拜占庭容错算法机制。

049

井通：构建区块链底层技术的标准协议

与比特币、以太坊和超级账本等以建设单一功能链（专业链）为主不同，井通致力于构建区块链底层技术的标准协议，努力打造可以承载多层功能的信用栈。

井通底层系统按照分层架构设计，将信任栈分为五层，实现了信任的分层传导；基于分层跨链技术，简化了智能合约的使用流程，便于实际操作；并在此架构的基础上提供了分级账本的功能，是目前国际上唯一实现分层架构的底层技术。

第一层是网络层，实现了点对点去中心化的数据传输，建立了数据传输的信任。

第二层是区块层，实现了区块链的基本功能，是不可篡改的且有时间性的区块构造所在，并以区块形式记录所有交易信息。这一层次是所有区块链的必备功能，并通系统将其功能独立，一方面是为了便于模块化，另一方面是已经可以满足一些基础性应用的需求。与传统的关系型数据库和 NoSQL 相比，它更适合于构建基于大批量用户数据一致性的应用，包括征信、溯源、防伪、数字版权等。

第三层是数据层，基于区块链层并允许所有节点使用自己的数据库技术将信息写入区块，实现了大数据和区块链两种技术的融合。

第四层是价值层，实现了数字资产生命周期管理，以及资产的发布、交易、互联、交换、冻结和授权等功能。基于我们对区块链是价值互联网的理解，价值层提供了多种价值共享 API。

第五层是合约层，实现了并通系统的合约功能。这一层提供智能合约功能，并基于价值层实现复杂的商业逻辑计算功能。

并通认为，区块链的发展成熟是多层次的发展成熟，而不是一个单一的成熟路径。分层设计的区块链中，每个层次的成熟度不一样，可以根据每个层次的成熟情况推出相应的区块链产品，从而增加业务的灵活度和层次感。

2.1.4 区块链技术的业务内涵

如果用一句话来介绍区块链技术的业务内涵，就是通过共识机制、密码技术和分布式多账本，形成多方共享的、不可篡改的、具有时间戳的信息链和交易链。

从诞生至今，区块链技术已经历了三代演变。

第一代区块链技术中的比特币底层技术，其核心就是解决了陌生参与方

的多重信任问题，发展出了去中心化、不依赖第三方认证的防止多重支付的技术解决方案，大幅降低了中间交易和支付费用。

从技术上来说，区块链都可以做到完全去中心化，这也是第一代区块链技术的鲜明特点。区块链上的验证机可以任意加入，不受限制。每台验证机享有完全平等的权力，共同参与全网的工作验证和收益获得的分配，其主要特点是机机平等。

第一代区块链技术中的比特币、以太坊是公有链的常见技术表现形态。完全去中心化的目标是绝对的自由、绝对的隐私，但从现实来看，这只是一个理想。纯粹的自由不一定好，全民的选择也不一定是最优的。

但是，第一代区块链技术存在明显的效率低下的问题。例如，比特币的任何交易和支付至少需要 10 分钟才能初步完成（通过改进的其他系统可以缩短到 2 ~ 3 分钟），60 分钟才可以最终确认，无法满足很多业务对即时性的要求。

第二代区块链从技术上提高了运行效率，开始考虑用经过选择的节点作为验证机来形成系统共识。这个变化明显会带来效率的提高，交易支付的时间可以缩短到几秒。

第三代区块链技术的核心是变“完全去中心化”为“有效去中心化”，也就是区块链上的验证机受管理机构的限制，只有经过授权的合格节点才能成为验证机，才能享有同样的权益。其主要特点是以相对的平等换取一定的效率，优点是效率较高、更易商业化；主要代表是 Ripple、R3、井通，技术表现形态多为私有链、联盟链。

有效去中心化有诸多益处。首先，不再需要大量“挖矿机”，既可以节约成本，也可以避免资源和能源的浪费。其次，发展方向上做到先实现行业内互联，再扩大到跨行业互联，最后实现万物互联，其循序渐进的节奏更具有

商业可行性。另外，有效去中心化还可以利用选择性的决策来避免群体性的失误。在第二代区块链技术版本上发展的第三代区块链技术，首先是改动底层的区块链，然后在这个层次上面建立了实名认证的用户体系。这个用户体系和区块链信息的关联是属于非公开的保密信息，用来保护企业和个人的隐私。当然，国家监管部门出于对金融安全、反洗钱和反恐金融的原因可以随需查阅，并可以提供限制、封锁账号和监控的功能。最后就是在用户体系上建立类似于社交网络的关联逻辑，提供低成本的信用体系评估。

无论是完全去中心化，还是有效去中心化，这两条路都有它的市场和拥趸，可以充分竞争，更可以各展所长、取长补短。

区块链技术具有三个明显的特性：公开性、安全性和唯一性。

公开性是指区块链中存储的信息对所有的参与者是完全公开的。这一点是由区块链的点对点网络（Peer to Peer）存储方式决定的。在区块链网络中，每一个节点都可以存储区块链的副本，而区块链的唯一性能保证这个副本在不同节点之间是完全相同的。

安全性是指区块链上的信息是通过数字加密技术保存的，只有掌握相应解密信息（私钥）的成员才能够解读信息。其他成员虽然可以看到并验证信息的完整性和唯一性，但无法获得私钥本身。

唯一性是指区块链上存储的信息是不可更改的。这既包括在空间上的唯一性，即所有节点都只有一个相同版本的信息，也包括时间上的唯一性，即历史数据不可更改。同时，区块链的唯一性还指区块链在运行过程中保持唯一一个链条的特性。因为如果出现不同的链条，区块链就形成了分叉。分叉的出现会使区块链在两个不同的空间维度中出现副本，而这正是要通过共识的规则来避免的。

2.2 区块链的商业价值

2.2.1 价值共享世界的 TCP/IP

如果我们回过头来看互联网技术的发展历史和网络构成，再对比现在的区块链技术和越来越热的价值互联的概念，可以发现区块链技术确实有可能成为未来价值互联网的 TCP/IP，如图 2-3 所示。



图 2-3 区块链是价值互联网的 TCP/IP

1968 年，美国国防部高级研究计划局组建了一个计算机网，名为 ARPANET(Advanced Research Projects Agency Network，又称“阿帕”网)。时逢美苏冷战，美国国防部认为，如果仅有一个集中的军事指挥中心，万一被苏联摧毁，全国的军事指挥将处于瘫痪状态，所以需要设计一个分散的指挥系统。它由一个个分散的指挥点组成，当部分指挥点被摧毁后，其他点仍能正常工作，而这些分散的点又能通过某种形式的通信网取得联系。发展到 1975 年，全球已有大量新的网络出现。由于最初的通信协议对节点以及用户

机数量的限制，建立一种能保证计算机之间进行通信的标准规范（即通信协议）显得尤为重要。1983年1月1日，所有连入ARPANET的主机实现了从NCP向TCP/IP（传输控制协议/网际协议）的转换。为了将这些网络连接起来，美国人温顿·瑟夫（Vinton Cerf）提出一个想法：在每个网络内部各自使用自己的通信协议，在和其他网络通信时使用TCP/IP。这个设想最终导致了Internet的诞生，并确立了TCP/IP在网络互联方面不可动摇的地位，基于TCP/IP的网络推动了互联网的发展。

互联网在发展的过程中具有几个比较重要的特征。

（1）开放性：新的节点只要接受标准协议就可以接入网。

（2）去中心：互联网产生的初衷就是防止单点中心被摧毁而采用的多中心系统，随着后期更多的局域网和新组网技术的加入，互联网去中心的属性越来越明显。

（3）局域网（多样性）+ 互联网（一致性，TCP/IP）：互联网展现出一定的多样性，能够包容各类技术、协议和网络。同时，互联网又具有一致性，网络之间主要的通信都可以通过标准的TCP/IP解决。

（4）协议分层：互联网兼具多样性和一致性的源头，就是协议分层的实现。无论是五层网络模型，还是七层网络模型，核心思想都是在保持底层一致的前提下根据具体的应用扩展出市场需要的多样性。

（5）从量变到质变：互联网的发展是一个从量变到质变的过程。当接入的节点数量相对较少时，网络上的应用数量和社会对网络的关注也就相对要少，整体处在技术积累的阶段。当接入的节点达到一定数量后就出现了应用的爆发，也就是所谓的互联网革命。

回顾了互联网发展的历史，我们重新聚焦在区块链，把区块链技术放到

通过交换创造价值这个大概念中，去发现价值共享世界形成过程中的机会。

人类社会发展过程中，一直在追求更方便、快捷的价值交换。按照交易方式，人类社会可以大致地分成三个阶段：物物交换、一般等价物（货币）、信用经济。

互联网对世界上的信息流通产生了根本性的变革。在传统的模型中，每传递一次物流和资金流，约有 3 倍的信息流产生和传递。互联网接入后，可以在同样的时间传递 100 倍以上的信息流。信息的充分沟通带来了物流和资金流的优化。但是，互联网本身解决不了信任的问题，即使可以高效地实现信息和有价凭证的传输，依然要引入第三方背书解决互信的问题。

随着区块链的产生，第一次用技术手段解决了交易中的信任问题，第一次用技术实现了为交易背书。互联网与区块链的结合，可以产生类似互联网和共识协议的效果。

首先，互联网提供了一个无所不在的联通网络；其次，资产和价值两端提供方（如银行）的 IT 系统和公司的 ERP 系统可以看成是一种基于价值交换的局域网；最重要的是当区块链技术演变成建设基于类似 TCP/IP 之上的一种共识协议后，我们可以借助网络连接各个局域网，构建出能够在全世界范围内进行资产交换和价值交易的价值互联网。

价值互联的核心是要实现资产的互联互通和自由交易。要把全社会的资产都搬到价值互联网上，必须先解决三个问题：完全开放的平台、资金的安全和资产的定价。

区块链技术正好可以给出一个比较完美的解决方案——去中心化的网络。商业巨头们在保证自身数据安全的前提下与合作伙伴进行数据的互联，甚至共享；加密技术足以保障资产的存放和支付时的可靠、便捷以及安全；快速

的验证可以比传统系统更快、更安全地完成支付、确权以及资产的交割。

加入这样一个价值互联网，会比传统的中心化系统实现互联来得更加简单和方便。传统中心化系统要实现互联，只能一个系统一个系统地打通。尽管标准化和信息服务技术的进步使打通的过程不那么痛苦，但是它们并没有带来革命性的变化。而新的思路则是把要打通的数据放到区块链上即可。根据实际的需要，这个区块链可以是私链（自己公司内部共享），也可以是联盟链（集团内或者合作伙伴间共享），更可以是公链（全社会共享）。

在区块链网络中，商家们可以共享自己的资源（商品或服务、客户资源、支付手段），普通用户们也可以在网络中共享自己的资源（如闲置的资产）。尽管资源方有强有弱，但是不会出现绝对的垄断者，更不会出现强者通吃的局面，这与政府所倡导的互享互通的共享经济不谋而合。

区块链用自己的技术优势保证了价值互联网络的信息公开、透明、准确，并保护了参与者的隐私和数据安全。这样一个以全部互联网为基础的价值互联平台包括了未来广泛的社会资源和社会资讯数据，为未来社会化大数据时代的到来奠定了坚实的基础，做好了充分的技术准备。

未来的某一天，基于区块链技术的价值互联网将作为生产关系的一部分，对社会生产力产生重大影响。因此，从宏观意义上讲，区块链技术是一种可能会对人类社会产生重大变革的技术。

2.2.2 全社会智能资产化

区块链可以用于任何数字资产的注册、存储和交易，包括金融、经济、货币等各个领域的有形资产（物理资产）及无形资产（投票、概念、信誉、想法、健康数据和信息），为各种不同类型和层次的行业应用提供连接与

融合。

智能资产一般是指一切以区块链模型为基础的可交易的所有权类型资产。这些资产可能是物理世界中真实存在的资产，如房屋、汽车、自行车或计算机，也可能是类似于股票、储蓄或版权（如书籍、音乐、绘画以及数字艺术）这样的无形资产。任何资产都可以在区块链中注册，其所有权是被控制私钥的人所掌握。所有者能够通过转移私钥或者资产给另一方来完成出售资产的行为。

广泛构建无需第三方信任的去中心化资产管理系统，以及通过加密的方式来控制资产，这也许能够在物权法律的范围内获得极大的应用，通过对资产本身进行记录来极大简化资产所有权的管理。

自动执行的智能合约代码被绑定写入代码底层并无法剥离，这就无法阻止编码预先设定的产权交易的发生。智能合约的特点是双方之间同意做或者不同意做某事，但无需再信任彼此。事实上，智能合约之所以能如此操作主要是有三个要素：自治、自足和去中心化。首先，自治表示合约一旦被启动就自动执行，而不需要它的发起者进行任何干预；其次，智能合约能够自足地获取资源，也就是说通过提供服务或者发行资产来获取资金，当需要的时候也会使用这些资金；最后，智能合约是去中心化的，也就是说它们并不依赖单个中心化的服务器而分布式存在，并且通过网络节点来自动运行。

智能资产和相关智能合约也需要符合现行的法律规定。例如，预先建立的智能合约能够在某个人已经偿还全部贷款后，自动将车辆所有权从财务公司转让到个人名下（还需要其他以区块链为基础的智能合约来配合）。同样在未来的某一天，抵押贷款利率可以通过调用另一个以区块链为基础的智能合

约以及有合约数据的网站或数据接口而被自动重置和修正。

区块链技术提供一个重塑身份认证和安全进入的方式。它有更细的粒度，能更灵活地面向实时需求，远比当前的技术选择有更多的可行性。而且，区块链技术可以通过嵌入物理世界硬件中的方式被方便地使用。

2.2.3 对接机器经济

物联网（Internet of Things, IoT）是建立在互联网之上的，是把所有事物联系在一起的网络。IoT 中的“Things”是指一切事物，可以是任何一个普通物理对象。高德纳公司预测到 2020 年，物联网设备将达到 260 亿台，能够创造 1.9 万亿美元的产值²。而且，随着机器智能化程度的提高，物联网设备的数量及创造产值的增长速度将会越来越快，这样就需要一个微支付的网络来连接并管理智能设备间的交易。

未来可能出现的场景是：自动售货机卖完商品之后通知无人送货车补货；当补货完成之后，自动售货机自动把货款支付给无人送货车。区块链技术可以被完美地应用到这个场景中。因为机器间无法处理信任的问题，区块链技术带来的去第三方信任机制正好可以弥补机器间无信任关系的这块短板。

这一切并不是纯粹的空想。IBM 和三星最近提出了一个应用设想，让家用电器如洗碗机可以通过执行智能合约来发布命令，要求洗涤剂供应商进行供货。智能合约给予了设备支付订单的能力，还能够接收来自零售商的支付确认消息和发货消息，并以手机铃声提醒的方式通知物主。

² 高德纳在 2013 年 12 月 12 日的报告中称，“到 2020 年，物联网设备数据将达 260 亿。”

在去中心化的物联网中，区块链是能够促进交易处理和交互设备之间协作的基础架构。每个区块链管理自己的行为，发挥自身的作用，这样就会形成一个“去中心化的自治物联网”。未来的经济行为可能不再局限于人与人或者组织之间，机器之间也会发生经济行为，区块链技术将开启机器经济的时代。

未来的物联网将会是大数据的一个重要来源。如何提高物联网大数据的质量，如何有效筛选、整合、处理物联网大数据，区块链有能力发挥巨大的作用。

现在的物联网依旧是被动运作和产生数据，未来更进一步的机器经济领域会是人工智能。人工智能超越物联网的地方是可以进行主动控制和信息筛选，对大数据来讲是极大地优化了基础数据的产生，而不是像物联网传感器那样不加区分地采集数据。可以说，人工智能就是物联网的升级换代。或者说，传统物联网的相当一部分迟早要被人工智能所取代。可以预期的是，未来的人工智能会成为大数据的一个重要数据来源。

但是，人工智能高质量的信息和服务更加需要通过区块链技术搭建的价值互联网络来传递自身产生的价值，而且从技术上也将会更加适应区块链网络，甚至可以用独立身份参与价值互联网络的各种商业运作。人工智能产生、掌控的大数据和其他智能资产会如何在区块链价值互联网络上传播和交易，前景可能会远远超出今天人类的想象。

2.2.4 优化社会构成

区块链的本质是一个大规模协作工具，其影响绝不局限于个别地方，而是会改变我们所有人的协作方式。

著名社会学家凯文·凯利在他的著作《失控：全人类的最终命运和结局》

一书中论述了人类社会及科学技术将如何进化。他把工业社会的进化论总结为基于机械逻辑的进化论，把信息社会的进化论总结为基于生物逻辑的进化论。这种基于生物逻辑的进化论，概括起来可以用三个词描述：分布式、去中心、自组织。

一个怀着蜂群思维³的群体，相比传统的中心化组织有着可适性、可进化、弹性、无限性、新颖性等多方面的优势。这个思想在技术上的应用，使我们在通信网络等技术领域取得了巨大的进步。借助区块链技术，我们可以在真实的社会组织中实践这些思想。去中心化自治组织（Decentralized Autonomous Organization, DAO）、去中心化自治公司（Decentralized Autonomous Corporation, DAC）和去中心化自治社会（Decentralized Autonomous Society, DAS）在很大程度上可以解决相当多的实际问题。

人们在作决定的过程中，尤其是涉及一些公共决策时，投票仍然是解决问题的常见方法之一，它为每个人提供平等的机会。尽管世界上各个地方都会使用电子投票系统，但是仍旧需要花费几个小时来进行人工验证。即使在那些有长期投票传统的国家，舞弊的可能性也依然存在。而如果把整个投票过程都记录在区块链的分布式账本上，其不可篡改、真实可信的特性将会使投票结果更具信服力。

我们可以通过投票对具体的事务进行决策，但是仅凭决策而没有自己的资产还不能支持一个组织的运营。DAO 就是构建在去中心化平台之上自主运

3 蜂群思维是一种集体思维，是指由许多独立的单元高度连接而成的一个活系统。它的神奇在于没有一只蜜蜂控制蜂群，但是有一只看不见的手控制着整个群体。它的神奇还在于量变引起质变。要想从单只蜜蜂的机体过渡到集群机体，只要增加蜜蜂的数量，使大量蜜蜂聚集在一起，使它们能够相互交流。等到某个阶段，当复杂度达到某一程度时，集群就会从蜜蜂中涌现出来。蜜蜂的固有属性就蕴含了集群，蕴含了这种神奇。

行的、拥有内部资产的实体。

严格地说，DAC 是 DAO 的一种形式。这个术语最早是由丹尼尔·拉里默（Daniel Larimer）提出的，他一贯坚持的一个观点是 DAC 要支付红利。也就是说，DAC 中有股票的概念，并可以某种方式进行购买和交易，这些股票让持有者持续享受 DAC 的成功。随着 DAO 和 DAC 的成熟，整个社会最终将会形成 DAS，人类文明将进入一个新的阶段。

区块链社会就是 DAS 的一种形式，它是指通过区块链技术，以去中心化、更便捷、更有效和更个性化的方式来实现政府和社会管理机构所提供的传统服务，以及更多的、全新的社会管理和服。政府和社会机构能够利用区块链技术的优势，公开保存社会档案，方便查询和进一步的社会管理。

061

DAS 的另一层重要含义是个性化，政府管理和服务可以从统一标准的单一模式到根据个人的需求进行定制。例如，某位居民可能要为更高级的垃圾处理服务支付更多的费用，而他的邻居需要为更好的学校来付费。“区块链上的政府”是有可能更加真实地代表民意，更敏锐地理解民意需求，更灵巧、敏捷地作出政策和规定调整，优化社会结构，改善社会治理，促进社会进步。DAS 的出现为政府提出了如何优化社会治理结构的思考。人类权利和义务的构架将如何变化，形成怎样的治理结构，这些都令人十分憧憬。

DAO、DAC 和 DAS 这些概念会随着区块链的推广而越来越多地受到有识之士的推崇，科技创新的方式必然由传统的工程师团队作业逐步向个体化过渡。在一个成熟的分布式自治系统内，没有“权威”这个概念，任何人都可以不受干扰地向其他人表达自己的观点和理念。只要方案足够巧妙、逻辑足够缜密，就可以取得足够的支持，并且所有服务将会是全球化的。

那么,在这个过程中,政府管理和社会运行将会全面地大数据化,社会化大数据将应运而生,并成为人类生活必不可少的组成部分,区块链技术将会是其中最重要的推动力量。

清华大学出版社

北京 100084

第 3 章

对技术的哲学思考

世界观是人们观察和思考世界运行规则的底层逻辑，因而隐藏在我们思维的背后。世界观决定了我们思考的方法论，决定我们的现在和未来。所以，在新知识、新发现、新科学不断涌现的现在，在研究大数据和区块链的现实价值与未来影响时，我们试图从世界观底层做出些许梳理，从哲学视角做出长焦审视，或许是很有价值的。

技术从来都不是无缘无故产生的，也绝不是孤立存在的。新技术、新思维的探索就像时代的脉搏，永不停歇，永无止境，既凝聚着对未来的无限希冀，也汇聚了历史的厚重积淀。本章将会追寻世界观的脚步，探寻区块链和大数据技术真正的价值和未来的发展。

区块链和大数据技术也是伴随人类社会一步步共生进化而来的。然而，无论人们是否关注过这些深邃的底层逻辑，构筑怎样的世界观体系，这一切其实都是表象，现实的驱动力量只有一个，那就是效率！因此，本章还会着重探讨效率背后的脉络，不仅是探究思想，更希望能照见未来。

3.1 大数据与不确定性

面对汹涌而来的数字化浪潮，早在 2008 年，《连线》杂志主编安德森就认为，大数据会深刻地改变人类探索世界的方法，甚至认为量子力学脱离了实际，理论已经终结，只要有大量数据，将应用数学作为工具就可以探索世界。安德森的这个判断在大数据会带来重要影响方面没有错，真正错误的是在此言论背后，他对世界的认知仍停留在牛顿世界观，正像大部分人一样。

3.1.1 被牛顿刻画的世界观

065

在文明的初期，人类凭借自身的生活经验认为：大地是平的，周围有海洋，大地漂浮在海洋之上；太阳每天从东面升起，每晚就在西边山后休息；地球不可能是圆的，否则海水早已经从“下面”流走；在地球的“下面”（另一侧）不可能有人，因为人不可能头朝下，还能站在地下方。

公元前 300 年前后，亚里士多德开启了对世界的科学认知。他认为地球是圆的，地表的曲率一致。他还认为地球是静止的，否则我们抛向空中的物体就不会仍落回我们手中。所以，地球是宇宙的中心，其他星体都围绕地球转动，宇宙是完美的。亚里士多德的《天论》和托勒密的《天文学大成》中对地球中心说进行了严密论证，形成了完整的科学范式。

世人对世界认知的调整是非常艰难的。直到亚里士多德死后近两千年的 15 世纪，地球中心说仍未完全占领人们的认知，仍有人否定地球是圆的。直到麦哲伦完成著名的环球航行，人们对大地是球体的怀疑才最终消除，地球

中心说逐渐成为主流的世界观。虽然用现在的眼光看地球中心说，许多观点是非常愚蠢和可笑的，但基于当时的科技条件，地球中心说的确是对世界最完美的理解和解释。

16 世纪，哥白尼发展了以太阳为中心的天文系统，不仅把太阳当作行星运行的中心，也把太阳当作整个宇宙的中心。太阳中心说和地球中心说是完全相反的世界观，本应对原有的世界观带来巨大的冲击。但是，哥白尼的天文学体系和托勒密的很多方面都一样，主要差别只是把地球和太阳的位置互换了一下，虽然在对世界运行的描述、预测和个别异常现象的探究方面有了一些更好的解释，但反而在一些重要的领域，如有关地球静止或运动问题，与当时条件下的物理学理论兼容性更差。所以，当时的科学家都熟悉但很少接受哥白尼的学说。

后来，开普勒以第谷的大量观测数据为基础，以火星为突破口，发现了更多能证明太阳中心说的证据。随后，伽利略用望远镜发现了更多的星体，突破了原有的宇宙范围，获得越来越多的实证。新的证据让地球中心说增加了更多瑕疵，这些证据很重要，也很有吸引力，但并不能根本性地直接解决两个学说的争论。太阳中心说因为在预测方面的一些优势，更多地只是被作为工具而不是世界观来应用。在两个多世纪里，地球中心说一点点坍塌，但想要彻底摧毁它，需要更清晰、更革命性的理论出现。这又一次证明，人类世界观的调整是异常艰难的。

直到《自然哲学的数学基本原理》的发表，牛顿用力学三定律和万有引力简单清晰地解释了世界的运行规律，才真正开启了现代科学的大门。

用一台高度精密运行的钟表或者机器来形容牛顿世界观，再也恰当不过。宇宙是巨大的，甚至是无限的，太阳是太阳系中行星环绕的中心。天体的运

动是由相互之间的作用力所驱动，没有目标或者目的，并严格按照一定的规律和轨迹运行，天体的状态和位置是可以推导和预测的。

世界是机械的、绝对的，运动是可以被精确计算的。事物是因果相关的，万物皆有因，万事皆有果。因果相应，因导出相应的果，因果连绵不断，故而未来的果是可以被预测的。同理，基于果可以导出因，也就可以向前推得最初的唯一因。因此，世界是连续的，基于因果驱动，事物是线性发展的，发展轨迹是连续的。

3.1.2 急需打补丁的世界观

18—20 世纪，在牛顿世界观的指引下，人类开启了现代科学的所有发端，取得了令人瞩目的成就。时至今日，现代科学的普适性教育都是以牛顿世界观作为底层逻辑，几乎所有现代人都接受了牛顿世界观的科学启蒙，牛顿世界观也深深嵌入了我们的思维方式。

在 1900 年伦敦举办的万国博览会上，开尔文男爵在回顾物理学所取得的伟大成就时说，物理大厦已经落成，所剩的只是一些修饰工作。同时，他在展望 20 世纪物理学前景时却若有所思地讲道：“动力理论肯定了热和光是运动的两种方式，现在，它美丽而晴朗的天空却被两朵乌云¹笼罩了。”

而正是这两朵乌云引领科学工作者在未知的领域探索，之后爱因斯坦的相对论以及量子理论就是两朵乌云结出的最亮丽的科学果实。刚刚建立了统治地位的牛顿世界观就需要进行调整，这个调整的过程也注定是异常艰难的。

1 两朵乌云：第一朵乌云出现在光的波动理论上，主要是指迈克尔逊-莫雷实验结果和以太漂移说相矛盾；第二朵乌云主要是指热学中的能量均分定则在气体比热以及势辐射能谱的理论解释中得出与实验不等的结果，其中尤以黑体辐射理论出现的“紫外灾难”最为突出。

► 相对论的启示

爱因斯坦发明了相对论，颠覆了我们许多常识性认知，以及许多明显且得到证明的事实。

狭义相对论有两个基本原理：相对性原理和光速恒常性原理。以前，我们认为时间和空间是绝对的，你 1.80 米的身高不可能是变动的，不管是在家里，还是在飞机上；流淌着的时间也是一样，你的一小时和嫦娥的一小时没有区别。但相对论告诉我们，时空是相对的，相对于不同的参照系是变动的。连时间和空间都不是绝对的，这比之前“地球是圆的”对“上与下是绝对的”的颠覆更令人难以置信。

广义相对论有两个基本原理：广义协变性原理和等效性原理。广义协变性是相对性原则的扩大版，讲的是物理定律在所有参照系中都是一样的，在所有参照系中相同的实验会有相同的结果，基于不同参照系进行观察，则结果是相对的。等效性原理是指加速度产生的效果与重力产生的效果没有区别，就是以相同加速度朝向地球的感觉和朝向任何方向的感觉没有任何区别。广义相对论的实质是研究时间、空间和质量之间的关系，爱因斯坦要用方程式把这个关系简单、有效地表达出来。

广义相对论对牛顿理论的一个重大冲击就是有关重力的学说。大多数受过牛顿世界观教育的人出于现实主义的态度，坚信重力是存在的。但事实是重力这种远距离作用力一直困扰着牛顿，莱布尼兹就批评牛顿将一个神秘的力量引入了科学，所以，牛顿对重力采用的是工具主义态度，因为重力可以完美解释和预测。

而广义相对论认为，质量会对时空产生影响，大质量的天体（如太阳）会造成周边的时空弯曲。在弯曲的时空中，直线不是直的，而是所在位置已

经弯曲了的“直线”。按照牛顿的惯性定律和广义相对论的运动定律，不受外力影响的物体会沿着直线（最短的）路线运动。在太阳周围弯曲的时空中，地球围绕太阳运动，走的正是“直线”。所以，太阳和地球之间并不存在所谓相互吸引的力。因此，地球上也并不存在什么重力。

牛顿的惯性定律说，运动的物体会保持直线运动，静止的物体会保持静止，除非受到外力的影响。惯性原理和日常经验完全是相悖的，因为人们日常感知到的是运动的物体总趋向于停下来。因此，惯性原理被称为17世纪最难发现的原理之一。

相对于惯性原理，接受爱因斯坦的相对论需要我们颠覆更多的常识。相对论的内容远不在我们生活可以体验的范围，弯曲的时空更超过普通人的想象。对于大多数人而言，牛顿世界观已经可以完美地解释世界。相对论虽然颠覆了牛顿世界观中的一些基础认识，但我们可以把牛顿定律当作相对论的特例，或者把相对论当作前者的延展。相对论的提出虽然已经有一个多世纪，但是与生活无关，更像书本上的知识，所以似乎并没有影响到普通大众的认知。

相对论告诉我们，世界是相对的，因为连时间和空间都是相对的，而不是绝对的；事物是与参照系相对应的，视角和参照不同，往往得出完全不同的结论。但相对论更重要的意义在于证明了我们会在常见的问题上犯错。这些常识往往被作为公识，想当然地、不假思索地用作其他推理的前提，因而必然导致许多荒谬的结果，也难以求得真知。

► 量子理论的启示

量子理论是研究微观物理世界规律的理论。微观领域有许多独特而深奥的特性，宏观领域的科学理论无法描述微观领域的规律。结合了牛顿经典力

学的量子理论就发展成了量子力学，结合了爱因斯坦相对论的量子力学建立了量子场论。从原子物理学、量子光学等学科的建立，再从半导体、晶体学到大规模集成电路等信息产业和现代新能源新材料的应用，量子理论的发展引发了一系列科学发明和技术应用。

量子理论的发展，结出丰硕应用果实的同时也必然带来巨大的认知冲击，我们的世界观也必然随之而变。但事实上，量子理论是极其复杂和难以理解的，连它的奠基人之一波尔都认为，“如果谁不为量子理论感到困惑，那就是他没有理解。”虽然我们的现代科学——工业领域到生活领域，几乎没有不依赖于量子理论的，虽然量子理论的诞生也已经超过一个世纪，但不可思议的是量子理论的基础思想和价值仍不为大众所熟知。

世界是不连续的。牛顿物理学理论认为，水温的升高和物体的运动是平滑和连续的。但是普朗克发现，能量的传递是不连续的，有一个最小的分量。历史上有个著名的芝诺悖论：一个人从 A 点走到 B 点，要先走完路程的 $1/2$ ，再走完剩下总路程的 $1/2$ ，再走完剩下的 $1/2$ ……如此循环下去，永远不能到达终点。按照这个逻辑，神话中最善跑的阿喀琉斯永远也不可能追上乌龟。《庄子》中也提到类似的悖论：“一尺之棰，日取其半，万世不竭。”这个悖论的假设是时空是连续的，在数学上是可以被无限分割的。但依照我们的经验，这个悖论显然是不成立的，时空不可能像数学计量一样无限地被分割，所以世界是不连续的。

世界是不确定的。科学家发现无法预测电子的行为，只能以电子出现的概率表达。经典物理学里面没有不确定这个概念，世界是机械的，一切都是可计算、可推导的；物理定律已对世界万物有所规定，事物的发展是注定和确定的。量子理论对确定性发起了挑战，随后混沌学又将其彻底打垮。洛伦

兹著名的“蝴蝶效应”表明，复杂系统中干扰是不可被预测的，任何微小的扰动都会带来难以预料的影响。决定论的失败是个好消息，给自由意志的生长带来了空间。

世界是非定域性的。包括爱因斯坦都认为，在一个地域的事件不能影响遥远地域的另一个事件，叫定域性。但是 EPR 实验、内尔定律和阿斯派克特实验发现，一个地方发生的事件可以瞬间影响另一个地方发生的事件，称为钟状影响，也就是事物的非定域性。例如，两个孪生的光源子，无论分开多么遥远都会仍像是一个整体一样瞬间影响到对方，并做出一致的协调动作。这在牛顿世界观看来简直是匪夷所思，这个影响犹如灵异般奇特。因为牛顿世界观认为，事物之间的相互作用是机械性的，是定域的，是需要直接关联的。钟状影响目前还仅是存在于电子等微观领域的研究中并得到证实，已经对牛顿世界观带来颠覆性冲击。如果有一天钟状影响扩展到宏观领域并得到验证，这个影响将是无法估量的。

071

从哥白尼的《天体运行论》到 1687 年牛顿发表的《自然哲学的数学原理》，牛顿世界观的形成经历了约 150 年。而从 1900 年普朗克创立量子论到狄拉克完成量子力学的普遍综合，量子理论的初步建立只花了 30 年时间。如今，量子理论已经是最实用的理论了。但迄今仍有许多未知等待探索，我们或许正处在一个科学探索的奇点，或许很快就会迎来更大的科学发现。

人类社会已经进化到 21 世纪，虽然我们都知道爱因斯坦，知道时空不是绝对的，宇宙是无限广袤的，量子理论也告诉我们世界是不连续和不确定的，但这些科技的发展远远超出人类大众的体验，主流的世界观仍停留在牛顿世界观时代。

在量子理论已经得到广泛应用的今天，我们更多是身在 21 世纪、脑在

18 世纪，因而迫切需要对世界观进行升级，或者是操作系统级别的迭代更换，或者是打补丁级别的版本更新。这个升级的必要性在于，要想充分理解当下的世界，理解现在的科技，理解未来的可能，都需要在当下最新世界观的基础上进行。

3.1.3 大数据是研究不确定性的技术

大数据技术中相关性代替因果性，混搭代替精确。维克托·迈尔-舍恩伯格在《大数据时代》一书中提到，我们没有必要非得知道现象背后的原因，而是要让数据自己发声。他还表示，相关关系能够帮助我们更好地了解这个世界，认为建立在相关关系分析法上的预测是大数据的核心。通过找到“关联物”并监控它，我们就能够预测未来。

在牛顿世界观里，世界是机械的，因果相应的；世界是确定的，是一台精密的机器；关系是精准的，一切都可以计量。而量子理论则表示世界是不连续的、不确定的，也是非定域的。

牛顿世界观带给我们树状思维，事物的发展就像根茎叶般因果脉络有序的；量子理论带给我们块状的块茎思维，这里没有因果、没有中心、没有层级、没有确定，只有相关、只有连接、只有流动，任何点都可以也应该建立连接，相关就应该连接。

我们以往的数据分析更喜欢依据图表上连续的线条作出分析和预测，这种线性的分析是小数据时代的一种惯性思维。这种惯性思维就是牛顿世界观里物体运动惯性的反应和运用。而大数据时代，事物的发展是不确定的，数据之间的关系是非线性、非因果的，毫无联系的事物之间也可能是关联和有

影响的，就像钟形影响的超距现象一样。所以，大数据技术的世界观底层是基于量子理论的。

基于不确定性的混沌学认为，一个随时间确定性变化或具有微弱随机性的变化系统称为动力系统；在这样的混沌系统中，初始条件十分微小的变化经过不断放大，对未来状态会造成极其巨大的差别。所以，混沌系统是不确定的和非线性的。混沌现象最先用于解释自然界，如“蝴蝶效应”对天气的影响，继而用于人文及社会领域，如股票市场、教育等。混沌学揭示了传统因果关系的错误认识。现在我们知道了，简单的原因可以产生复杂的结果，未来是难以预测的。

然而，探索因果和渴求确定性往往是我们无法遏制的内在冲动。因为一旦承认事物是不连续的，那么我们将失去预测的基础；一旦承认事物是不确定的，那么我们将会断线的因果链条中迷失方向。但是，类似钟形影响的相关性给我们带来了新的希望。

所以，我们为什么不可以这样认为，传统的因果关系将变化为更大范围事物背后更深刻的新的因果关系呢？如果可以的话，大数据技术有助于帮助我们消减不确定性，带领我们朝着新的因果关系进发。这种因果关系是基于相关关系之上更宏观的和弱化了了的新型因果关系。

相关就应该连接，伴随软件吞噬世界的是给任意连接的建立提供了可能性。新连接的建立会产生新的信息流动与新的能量交换，会催生新的价值和新的物种。相关性连接带来无限多的新可能，新连接也会促进新认知。基于大数据，认知计算开始吸收丰富的营养，快速地进化，并将赋予技术以思考能力。

3.2 区块链与信息基因

3.2.1 进化论

大约 135 亿年前，“大爆炸”形成了宇宙；大约 38 亿年前，在宇宙中的地球上分子结合在一起开始形成了有机体；大约 7 万年前，地球上开始出现了“智人”；后来，“智人”中的一支进化出了人类。我们是地球上数百万种生物中的一种，与其他生物是完全平等的。地球并非宇宙的中心，人类并非超自然的神奇力量。

世界是物质的，生命也不例外。构成生命的几十种化学元素，没有一种是生命所特有的，都可以从无机自然界中找到。但生命仍有独特性，生命是由蛋白质和核酸等物质组成的多分子物质体系，这些多分子物质是生命物质所特有的。而生命物质中 DNA 的自我复制是遗传、细胞增殖、生命繁殖的基础。

香蕉与人类共享了大约 60% 的基因，果蝇与人类共享了大约 61% 的基因，家牛与人类共享了大约 80% 的基因，老鼠与人类有着 85% 的基因相似度，阿比西尼亚家猫与人类有着 90% 的基因相似度，非洲猴与人类有着 96.66% 的基因相似度，黑猩猩与人类的基因相似度达到了 99.01%，我们每个人自己的基因与周围的人有着 99.9% 的相似度。

生命的最小单元是细胞，活的细胞能够从周围物质中吸取需要的成分，并生成自己所需的物质。当生长到一定程度时，细胞可以分裂成两个完全相同的细胞。细胞的分裂是从染色体开始的，染色体上最小的组成单元是基因，

其决定了生命的所有特征。基因本身是具有生命的“活分子”，每个基因大概由百万个原子构成。基因具有高度稳定性，因而可以保证每个物种具有稳定的特性。人体中基因的总量大约只有不到 60 克，就是这微不足道的几十克物质驱动了几千倍于自身的物质，构成了极其复杂的生命体。

目前发现的最小生命单元是烟草花叶病毒，大约由 200 万原子组成，它的繁殖过程与细胞分裂的过程完全相同。这种病毒正是连接生物与非生物的那一个“丢失的环节”，在生命起源中是关键性的一步跨越。

生物多样性、遗传变异和自然选择构成了进化论的主要内容。自然选择是进化论的最重要内容，是与环境相互作用的过程。进化中物种的特性没有好或者坏，没有高等或者低等，没有优或者劣，所有特性都是平等的。进化并非是目的性的，进化的过程并非要直接产生任何特定类型的特征，而是在会产生各种可能性的基础上碰巧产生了与环境更适应的特征，在生存竞争中取得了优势。所以，进化是高度不确定的。

075

自然选择又是被动和消极的。各种因素造成的基因变异为进化提供了有限的选择，进化无法突破选择的现实有限性。然而，环境却是动态和变化的，生存竞争也是动态和变化的，这在一定程度上是进化发挥主动性的领域。

3.2.2 共生进化

世界上有植物、动物、真菌、病毒、原生生物、原核动物六大类生物，数百万种，这些生物互为条件和环境，彼此交叉，彼此依赖，共生进化。我们胃里的微生物群落以我们的食物为原料，分解的产物才是人类可以吸收的营养成分；千鸟可以和谐地进入鳄鱼口腔，以鳄鱼口腔残留的食物为食，同时帮着清理寄生的水蛭；蜂鸟进化出鸟类唯一的特殊技能——在空中悬停和

后退，以便吸食花蜜，同时帮花朵传播花粉。

形式多样的共生关系是大自然中普遍存在的现象，也是生物进化的重要工具和特征。例如，树木需要发达的根系来吸收土壤深处的养分，但进化需要经历漫长的时间，有些树木甚至始终无法完成这样的进化，而土壤中的真菌恰好具有从土壤深处汲取养分的能力。当树木与真菌形成共生关系，显然要比单独进化更有竞争优势。

共生起源理论研究两种或多种共生生物的组合或关联而形成的生物体起源，将共生关系放到生命起源的原点来看待。现在看来是共生关系的组合，其实在生物起源时可能本身就是一体的。一个超级微生物群落或者活分子群落，内部分工协作，每一部分都好像是机器上的一个齿轮。其中一部分进化成了一种生命体，另外的部分形成其他生命体，之后仍共生进化。

共享、合作的共生对进化的贡献甚至超过了竞争。但是，这一点往往没有被充分地重视和理解。我们在学习和理解进化论时，一方面会看到生存竞争是自然选择的动力，物种获得相对优势会在竞争中生存下来，另一方面往往会忽视（过度繁殖导致的）生存资源有限性是引发竞争的前提条件，而且对基于共生关系的组合优势在进化中的普遍性不够重视。

正确的组合往往可以在竞争中取得更大的优势。在不确定的社会中，在块茎状的世界中，老的连接可能坍塌，新的连接随时可能建立；老的组合可能分裂，新的组合可能组建。创造性地连接、组合进化，既可能是创新的主要内容，也可能是创新的主要方式。

3.2.3 技术的进化

一切事物都在进化，世上不存在一成不变的东西。宇宙仍在继续生长，

每个天体都在经历自己的生命。我们生活的地球从最初的岩浆体，后来被其他星体带来的冰水浇凉了表面，变成了坚硬的岩体，岩体风化后变成土壤；地表被人类主宰后又出现大量的人造建筑。我们的思想每天都在变化，我们的企业每天都在经受激烈的竞争。

进化论适用于所有事物，不限于生物。如果说牛顿世界观是机械的、静止的、单向的世界观，那么进化论的世界观则是不确定的、变化的、发展的世界观。越是基础的越是稳定的，越是复杂的、高级的越是不稳定的，表现为更快速地变化。

技术当然也是进化的。如果把历史上所有型号的计算机、手机放到一起，这些型号的多样性会让人叹为观止。屏幕的大小、颜色、材质、设计等特性，让我们能明显地感受到这些产品是分“世代”的，这些不同的“世代”就好像生命体一样依次演进。从语言、符号、文字到书籍、收音机、电视机，从风车、蒸汽机到飞机、火箭，从大型计算机、微型机到笔记本电脑、智能手机，技术就像是第七种生命，仿佛有自己的进化规律。

《技术的本质》一书的作者布莱恩认为，技术好像都是由不同的基础组件构建的。所有的新技术都不会凭空产生，都依存于原有技术，脱胎于原有技术，遗传机制连接起了过去和现在；不同的技术组件在一起进行组合进化，不同技术组件的新组合是新技术产生的主要方式。技术的新组合就像是基因的变异，让不同的技术可以产生新的连接。形成的技术组合可以作为新的可遗传组件，作为未来更多、更新技术进化的基础。技术的高度积聚有利于技术组件的共享与新组合，为组合进化提供更好的环境，从而加速技术进化。

3.2.4 技术与人生共生进化

人类在驯化小麦之前是狩猎采集者，食物丰富多彩，生活悠闲自在。1万年前，小麦在被人类驯化之前仅仅是众多杂草中的一种。但仅过了1000年，小麦就突然传遍了全世界。小麦需要精心地料理，平整土地、除草、捉虫、浇水、守护，每一样都需要艰辛的劳动。人类在驯化小麦的同时，自身从狩猎采集者变成了从事农业的定居者，整个社会生活方式也发生了巨大改变。人类驯化小麦的过程也更像是小麦驯化了人类，是人和小麦的共生进化。

人类的需要驱动技术进化。某项技术于某个时期在不同地方会神奇地集中涌现，就是人类与技术组合进化的明证。技术是流动的，倾向于流向人类需要的地方。通常只有某个具体问题需要被解决，新技术才会出现。而这个新技术的出现又会引发其他难以预料的新技术。例如，人类需要更好地传播知识就有了印刷术；印刷术促进人类大量阅读，出现了视力问题，玻璃在望远镜中的应用技术就流动到了眼镜；之后镜片技术流动到研究微观世界领域并出现了显微镜，促进了对生物的研究。类似的还有计算能力从大型机流向小型机，流向个人电脑，流向更小的智能手机。

技术进化会逐步改变人类的属性。技术会放大人类的能力：衣服放大皮肤能力，望远镜放大视觉，电话放大听觉，汽车放大行走能力，外骨骼技术放大人体机能，计算机放大思考能力，手机放大连接世界的能力，这一切就像人类器官的技术外化。不仅如此，肾透析就是肾功能的替代，人工心脏就是心脏功能的替代，技术内化甚至开始替代人类的机能。医疗技术的进化会提高人类的存活能力，生命技术的探索会提高人类的寿命，干细胞技术可以复制人类的活体器官，自动驾驶技术代替人类的驾驶能力。

未来，人类与技术的关系是共生进化。就像生物界的共生一样，人与技术相互交叉、相互缠绕、相互依存、相互促进，共生进化。技术与人类有不同的需求，技术依附人类而进化，人类依靠技术而增强能力；技术与人类的界线在消融，技术拥抱生命，生命嵌入了技术，从而相融为一体；人类像对待孩子般培养技术，来延展能力和生命；技术与人类没有竞争，而是作为一个共生的整体参与竞争。

人类与技术共生进化的结果：人越来越依靠技术，越来越趋向技术，越来越像技术；技术与人共生进化的结果：技术越来越依附人类，越来越趋向人类，越来越像生命。人类与技术日趋统一，传统意义上的人类或许将不存在，代替的是更技术化的人类，更加智能。

3.2.5 作为信息基因的区块链

在与人类的共生进化中，技术想要拥抱生命，就会长入生命，嵌入人类的认知与行为并无法剥离，表现出技术自身的组织活性和生命力。技术越来越生命化，但真正赋予技术以生命的或许正是区块链。

我们先看一下这些概念的对应关系，如表 3-1 所示。

表 3-1 概念对照

传统人类世界	区块链世界
原子	0/1
分子	Bit
DNA	哈希算法
染色体	哈希指针
分裂	复制
细胞	智能合约
币值	能量
物种	应用

哈希算法²有几个重要的特性。第一是良好的碰撞阻力，以保证函数计算的唯一性。唯一性就是独特性，这个特性用在对信息的摘要。经过哈希算得出的统一规格的信息摘要，可以成为每个信息独特性和唯一性的有序表达。第二是隐秘性，即一个哈希值无法进行逆向，也就无法得到破坏和篡改，从而保证了信息的安全性和稳定性。稳定性是信息传递（遗传）的基本需要。第三是谜题友好性，说的是无法通过搜索的办法得到答案。在猜谜时，如果采用搜索的办法来获得答案就会面对一个巨大的空间，该空间远远超过当前的计算能力可以算尽的范畴。也就是说，这个巨大空间决定了现有计算能力条件下搜索的不可能性，同时给多样性提供了巨大空间。对应于 DNA，哈希算法的这些特性可以保证信息的唯一性、稳定性和多样性，这也正是 DNA 具有的基本特性。

哈希指针是一个指向存储位置和位置数据的哈希值，用哈希指针可以构建各种各样的数据结构。这个数据结构如果是二分状的，就被称为梅克尔树，从一个单一区块进行两两分化，就会得到一个层层递进的区块树。这个数据结构如果是链状的，就被称为链状区块链，上一个区块指针被置换为哈希指针，环环相扣，组成链状。就像染色体对 DNA 的有序组织，基于这种有序化的组织在繁殖中发挥重要的作用。哈希指针就是哈希值的有序组织方式，区块链可以形成一个无法篡改的记录，保障信息传递过程的稳定性。

2 哈希（Hash）算法即散列函数，可以将任意长度的二进制值映射为较短的固定长度的二进制值，这个短的二进制值称为哈希值。哈希值是一段数据唯一且极其紧凑的数值表示形式。如果一段明文哪怕只更改该段落的一个字母，随后的哈希都将产生不同的值。要找到散列为同一个值的两个不同输入，在计算上是不可能的，所以数据的哈希值可以检验数据的完整性。该算法一般用于快速查找和加密算法。

基于区块链的智能合约是能够自动执行合约条款的计算机程序。智能合约的实现需要事务处理和保存机制，以及一个完备的状态机。智能合约系统根据事件描述信息中包含的预设条件，自动触发输出预设的资源。智能合约的核心就在于以事务和事件的方式经过智能合约模块的处理，输出仍是一组事务和事件。

因此，这个通过机制就可以组织起各种功能组件且让组件实现一定行为功能的过程，就像由细胞组织在一起形成的器官能发挥一定的完整功能一般。不同功能组件（细胞）之间又可以被某个系统组织起来，形成不同的应用（物种）。功能组件是基础单位，存在无限的可能性，因而可以组合的应用也是无限可能的。基于功能组件的组合进化是应用进化的主要内容。

信息和能量在物质间的流动是生命和技术共同的物质特性。就像生命需要吞噬食物获取能量进行生长繁殖一样，区块链的运行也是需要消耗能量的，这个能量包括但不限于电力。对区块链中能量的价值赋值，智能合约就会让信息的流动更有目标性。这个目标性就像是生命的本源属性一样——吞噬、转化、生长、繁殖，其不是指技术依附的人类的目标，而是指信息自己的目标。这种目标是由信息组成的应用所无法控制的，会自动发生，就像人类无法知觉和控制自己的基因一样。

技术进化中的自然选择是基于技术与人类的共生关系而进行的。曾经存在独立的人类进化，也存在独立的技术进化。随着技术与人类共生趋向一体化，人类的进化就是技术化，技术的进化就是生命化。所以，技术和人类越来越形成一个进化整体。之前的技术进化只是看起来好像具有生命特征，而现在的技术进化就是真正具有了生命特征。

区块链就是赋予技术生命能力的开始。人类是碳基生命，是生化算法驱

动的生命，是基因生存的机器；技术是硅基生命，是电子算法驱动的生命，是区块链生存的机器。

基于不确定性和相关性的大数据，正在赋予技术以思考能力。基于去中心化、共享、共识、共担的区块链，正在赋予技术以生命能力。区块链和大数据一起赋予技术具有智能的生命能力，从此，技术和人类开始新的共生进化。

3.3 竞争力量的源泉

权利是否对等、能量交换是否均衡、信息流动是否均等，是判断一个系统是否中心化的标准。在中心化系统中，中心决定节点，节点必须依赖中心，离开了中心就无法生存。因为中心是能量的输出方，其权利大于节点，所以更多地承担发出信息的功能。而在去中心化的系统中，节点与节点之间的影响会通过网络形成非线性的因果关系，这种系统是开放式、扁平化、平等性的，任何节点都可以成为一个中心，任何中心都是阶段性的，并且对节点都不具有强制性。去中心化不是不要中心，而是由节点来自自由选择中心、自由决定以谁为中心。

3.3.1 中心化的力量

早期的智人体型不够大，力量也不足，在生存竞争中并不具有优势。散落各处的每个部落里也就几十号人，部落周边全是野生动物，每一种大型动物都可能伤害到人类。这时候的人类远没有到妄自尊大的程度，他们会充分

地尊重周边的邻居，努力地在与邻居的协商与竞争中生存。但是到了农业社会，人类感觉到自己的力量已经非常巨大。虽然种植作物也非常不易，需要看天吃饭，但驯服了牛、马、羊、猪来为自己服务，就已然开始不用考虑其他生物邻居的想法了。

人类区别于其他生物的最重要的一个能力是讲故事。我们通过讲故事来构建一个愿景，组织起大规模的群落，实现大规模的协作，从而展现了惊人的力量。从智人开始统治地球至今的7万年中，人类已经成为影响全球生态变化的最终因素。中心化的人类社会组织能更有效地利用资源，更有效地分工协作，使人类自己在生存竞争中居于绝对优势。

国家的出现是社会发展的必然结果，也是中心化竞争优势的终极体现。在一定地域中，人们自发地建立一个共同体，这就是国家。国家是一种制度性的权力运作机构，拥有治理社会的权力，在一定的领土内拥有外部和内部的主权。国家产生之后，人们在物质资料生产过程中结成的生产关系逐渐代替了血缘关系，使社会结构发生了根本变化。新的社会制度取代了由血缘关系决定的氏族制度，这就是具有公共权力的国家制度。

公司是另一种中心化力量的极致体现。基于使命和愿景，公司组织凝聚起个人，具有比任何个人都强大的能力和经济动力，使血缘、地缘联系之外的陌生人共同合作成为可能，并且能扩展一个经济单位所能掌握和支配的资源，分散商业活动的高风险。

公司无处不在，其所具备的力量已渗透到人们工作和生活的方方面面。全球化日渐加速的今天，数百家跨国公司正左右着世界经济的运行，显示了足以撼动世界的能量。从17世纪到20世纪70年代，被经济学家认为改变了人类生活的160种主要创新中，80%以上都是由公司完成的。今天，全世界

70%的专利和三分之二的研究开发经费出自跨国公司。公司的影响越来越大，甚至超过了国家。

中心化，通过分工协作来延展和放大个体的力量，展现了这种模式的竞争力；通过组织来筹划资源的总体利用效率，则表现出生存能力。所以，中心化是一种生存竞争哲学，它展现了巨大的力量。而这个力量表现为巨大的竞争优势，是促使万物走向中心化的驱动力。

数据时代已经来临，大数据正在赋予数据智能化。而中心化的数据具有规模化、多样性和复杂性的特点，从而得以发挥大数据技术的价值。因此，大数据成为加剧中心化的力量。中心化有利于信息积聚，而信息积聚会加速组合进化，加速了的进化趋向垄断。所以，世界最终会趋向有限的几个数字大脑。同时，技术开始接管人类的部分工作、部分职能，甚至部分机体。人类在与技术共生进化的过程中开始趋向技术化的超自然的“神人”。但并非所有人都自然会成为“神人”，只有那些掌握算法、掌握数据的人，能利用技术赋予的能力成为极少数的“神人”，才会成为世界中心的力量。“神人”会获得竞争中的优先权利、长寿长生的权利和资源分配的权利。

3.3.2 去中心化的力量

人类享受着中心化模式带来的高速发展，物质和精神不断富足，但伴随而来的却是自我异化之旅。

首先，随着对世界形成了更加科学的认识，人类开始从中心位置上逐渐滑落。人类在竞争中居于优势时就开始自我膨胀、自我神化，并忘记了历史，一切以自我为中心，甚至认为太阳也要围绕地球转。直到16世纪，太阳中心

论逐步奠定了基础，才让人类认识到自己生活在浩瀚无边的宇宙中，原来世界不是以我们为中心的。这种科学认知形成的过程是一步步还原真相的过程，把人类从过去生存竞争中产生的盲目自大的成就感中拉回现实。这个过程也是人类重新认知世界的过程。在牛顿建立了新世界观之后，技术给人类巨大的力量，人类才感觉重新掌握了世界，重新建立起了空前的自信。

但是，进化论很快就把人类又拉下了神坛。原来人类并没有得到特殊的眷顾，和其他动物一样只是万物生灵中的一种。甚至人类和其他生物一样都是碳基生命，都是由相同的物质材料组成的。生命的底层都是生化算法驱动的，甚至只是基因机器，由基因决定了生命形式、行为模式和生命周期。于是，人类又回到了物质的本源和生命的起点。

其次是人文主义的崛起。人文主义是一种世界观，是人类随着科学进步而信心高涨的产物。它肯定人性和人的价值，要求个性解放和自由平等，推崇人的感性经验和理性思维，并以个人的兴趣、价值观和尊严作为出发点，使人的自由意志得到彰显。因此，人文主义是一种去中心化的巨大力量。

个人不甘于也不应该只是一颗无关轻重的螺丝钉，而是真正有血有肉、有私心、有欲望、有价值、有尊严的。人文主义反对一切以神为本的旧观念，宣扬人是万物之本，用“人权”对抗“神权”；也必然会要求在国家政治生活中主张更多的个人权利，在经济活动中争取更多的个人利益。于是，人文主义开始影响到社会的方方面面，从政治经济到文化教育等，成为了基本共识和基础价值观。以开源运动为开端，以众智、众筹、共享为代表的自下而上的自制类趋势表现出强大的生命力。更多样化价值主张的社群开始空前繁荣，他们常常是以自制为动力，以去中心化为形式。

再就是互联网。互联网的起源在技术底层就是按照去中心化的模式进行

设计和运作的。互联网正在展示吞噬一切的能力，先是信息和社交，再到商业乃至万物。在这个被数字化的世界里，信息通过互联网可以高效率、低成本地流通，能力可以快速地交换，自由意志可以充分地表达，无需中心过多地管理和参与，世界变得更扁平、更公平。

在互联网中，每个个体都是一个节点，每个节点都可能或可以建立连接，每个节点都可以参与表达，输出影响力。互联网成为一种有力的武器，个人力量借助互联网得以被放大。过去需要通过公司或组织构建力量的模式，在互联网时代已经失效。更高效协作、更低成本和更凸显个人价值的虚拟组织或微组织已经显示出过去只有大型组织才具有的力量。基于互联网，个体可以自我赋权，如自媒体，同样可以形成巨大的力量去影响和改变世界。

最后，区块链或许会成为去中心化的核心力量。去中心化、自制、开放、透明正是区块链的底层逻辑，与基于不确定的量子世界观是完美匹配的，或许在未来可以更好地支撑去中心化的社会组织。作为信息基因的区块链赋予信息生命能力，搭建起有基础生命能力的信息组件，使信息组件的组合能进化产生丰富多样的应用。丰富的应用完全服务于自由意志，同时赋予个体丰富的能力。让个体在与中心化的“神人”的竞争中保持多种可能性，留存一定的希望，而不是必然地臣服于中心化力量。

3.3.3 无关道德，只看效率

一个系统参与生存竞争的结果是自身与环境动态地相互作用决定的，而不是事前能够规划的。但系统表现出的中心化与非中心化的特征是一种两极分化的逻辑抽象与总结，非此即彼。所以，在自然选择中胜出的系统总是可以分成中心化与非中心化的，可以看到这两种似乎在竞争的力量，只不过这

两者都是胜利者也同时是失败者。

有关中心化与否的讨论，一个显而易见的问题是将中心化置于道德评判的立场，似乎去中心化占据先天的道德高地。但与中心化相关的权力、集权、统治、封闭不应该是贬义词，就像与去中心化相关的分布式、自制、开放也不应该是褒义词一样。中心化与否，与道德无关，我们不应执迷于偏见。就像封闭式系统的苹果公司，可以高效地组织调动资源、高效地管理业务，一样可以缔造伟大的企业。

中心化与否是一个复杂而交织的概念，没有绝对的中心化与去中心化，一切都是相对而言的。在历史的长河中，某个特定范围内中心化占据主要地位。例如，人类在取得竞争优势统治地球时作为一个抽象的整体，相对于周边其他物种是趋向中心化的；而不同地域存在不同的国家，相对于人类的整体却是分布式的、去中心化的，是中心化与去中心化的混合体。又如交易所，它能集中处理信息，其中的个体参与者依据信息独立作出决策。从技术系统角度看，交易所是中心化的；而把个体的“人体芯片”的计算部分加入考虑，交易系统又是分布式的。

随着数据时代的来临，对互联网去中心化的看法也开始需要适当地调整，更准确的说法应该是去中介化。互联网环境下的超级应用模式、赢者通吃的规则显示出了中心化的力量；而这些应用又给个体赋能，显示出了去中心化的能力；大数据和云计算技术会导致后端越来越集中，用户需求越来越个性化，满足用户需求的产品会越来越多元化。因此，后端趋向集中化，前端趋向去中心的多元化，二者相辅相成。所以，互联网是通过去中介化的方式，同时加强了中心化和去中心化的力量。

作为逻辑抽象的总结，中心化与否并非事物背后的驱动力量。所以，中

心化与否，本身不能当作趋势或作为标准对事情作出评判。把中心化或者去中心化作为趋势或者力量的源泉进行系统地设计，以期待产生竞争优势的想法和做法，更是缘木求鱼。竞争力本身才是那个驱动自身被选择的力量，有竞争力才有机会在生存竞争中胜出，才有机会被自然选中。而且，竞争力可以体现在力量能力、组织效率、能量交换效率、信息处理效率等多个方面。

3.3.4 竞争力的内在实质

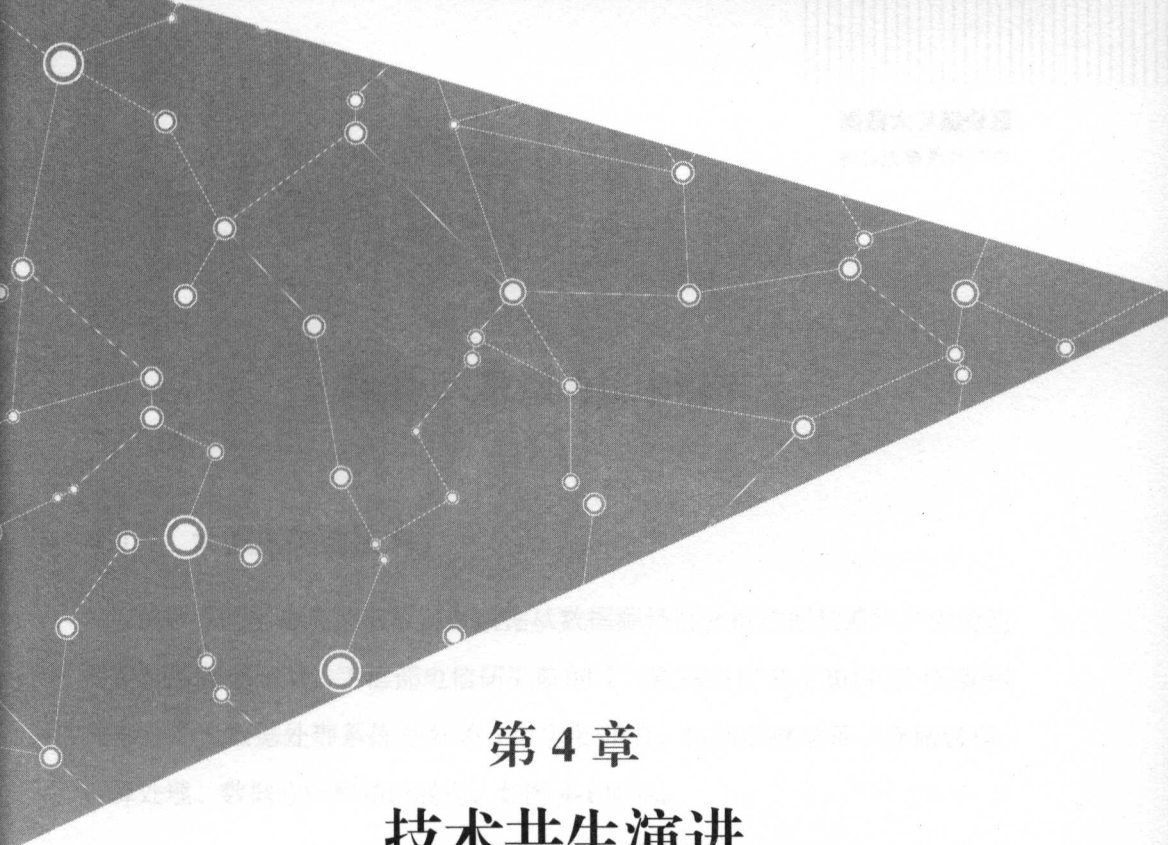
自人类进入农业文明开始，随着生产力规模的积聚扩展，中心化在人类生存斗争中具有更多的优势。因此，社会的组织结构和组织形式更多表现为中心化。但人类进入信息社会以来，新技术除了支撑中心化的效率提升，也为去中心化提供了技术支持。随着社会信息化连接的规模急剧增加，世界一体化特征日趋明显，世界更加扁平。未来，国家的定义会更丰富，社会组织形式也会呈现更多样化。当中心化在更大规模和更大范围的应用中居于统治地位时，中心化在人类生存竞争中的优势和价值或许已经达到了顶峰，去中心化则开始在一些领域中展现出巨大的优势。

去中心化和中心化竞争的核心在于效率，尤其在于连接的效率和流动的效率。蒸汽机、铁路、汽车、电话和移动通信都通过提升连接效率提升了世界运行的效率，人可以在更大范围更快速地进行移动，信息可以更高效地流动。例如，在支付领域，基于区块链的支付方式，其组织形式所产生的“摩擦力”更小，资金的流通效率更高，体现了明显的效率优势。这种优势在跨境支付中尤为明显。

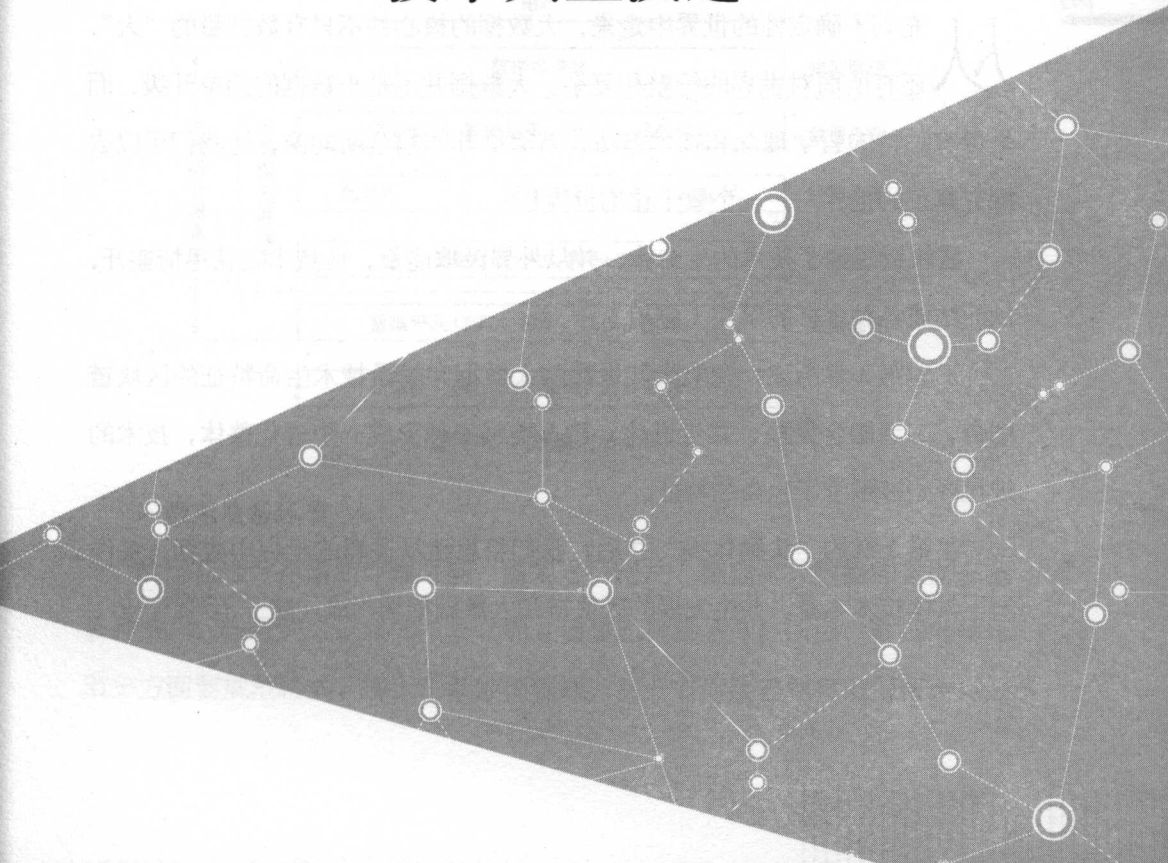
所以，去中心化不能作为区块链的优势，最多作为技术特点，这个特点可能在某些应用场景下恰好可以发挥出巨大的优势。同样，在匿名和隐私保

护方面的技术特点也不能作为区块链的优势，因为在操作层面很难割裂与现实身份的关联而做到完全匿名。面对区块链热潮，如何基于效率提升来思考应用才是问题的核心。

在分析有关效率的竞争问题时也要注意，在存量市场，具备一点点的优势不足以改变市场现有的格局。也许需要十倍或者更多才行，少量的效率提升难以冲破现有的惯性势能。换句话说，在存量市场，我们并不太需要仅具一点优势的技术和工具。类似哥白尼提出日心说，本身是颠覆性的或者貌似是颠覆性的，但与托勒密的地心学说仔细比较会发现好像相似性更多，而解释和预测优势并不明显，所以未能比托勒密的地心说带来更大的冲击。



第4章 技术共生演进



从充满不确定性的世界中走来，大数据的核心并不只有数据量的“大”，还有所面对世界的纷繁和复杂。大数据并不是小数据的简单升级，而是带有独特的哲学理念和技术方法，希望推开一扇全新的窗，让我们可以去探究真正的世界——一个量子化的世界！

区块链组建了技术的生命体，并从外界汲取能量，让技术之花尽情盛开，从而结成各种各样的果实（应用）。

正如第3章所述，探索未知世界的大数据和赋予技术生命特征的区块链结合，二者融合发展、共生进化，和人类越来越形成一个进化整体，技术的进化开始被赋予了生命的特征。

在第3章的“头脑体操”之后，我们将思绪从逻辑的世界中收回，关注一下具体技术本身。共生发展的区块链和大数据相辅相成，缺一不可。

4.1 大数据的技术沿革

4.1.1 大数据关键技术

从数据的生命周期来看，大数据从数据源经过分析挖掘到最终产生价值需要经历若干环节。工信部电信研究院的《大数据白皮书（2014年）》就将此部分的大数据处理系统划分为5个主要环节，包括数据准备、存储管理、计算处理、数据分析和知识展现，如图4-1所示。



图 4-1 大数据技术框架

► 数据准备环节

数据在进行存储和处理之前，需要进行清洗、整理。大数据的数据来源多种多样，数量庞大、格式不一，而且质量也良莠不齐。所以，数据准备环节一方面要规范格式，便于后续存储管理，另一方面要在尽可能保留原有语

义的情况下去芜存菁。

► 存储管理环节

如今，全球数据量高速增长。从可持续的角度来看，大数据存储系统不仅要以极低的成本存储不断增加的数据，还要适应多样化的数据管理需求，满足可扩展性。

► 计算处理环节

这个环节需要选择适当的算法模型快速处理数据。传统单机或并行计算技术在速度、可扩展性和成本上都难以适应大数据计算分析的新需求。分布式计算已经成为大数据的主流计算架构，但在一些特定场景下的实时性还需要大幅提升。

► 数据分析环节

从纷繁复杂的数据中发现规律并提取新的知识，是大数据价值挖掘的关键。传统数据挖掘一般会根据先验知识预先人工建立模型，然后依据既定模型进行分析。对于非结构化、多源异构的大数据集的分析，由于缺乏先验知识而难以建立显式的数学模型，分析者需要使用更加智能的数据挖掘技术。

► 知识展现环节

在嵌入多业务的闭环大数据应用中，一般是由机器根据算法直接应用分析结果而无需人工干预，知识展现环节不是必需的。而在服务于决策支撑场景下，如何让复杂的分析结果以直观的方式呈现给用户是主要挑战。

4.1.2 谷歌论文为大数据技术奠基

谷歌拥有先进的大数据管理和处理技术，能解决海量数据存储和数据处理问题，因而在搜索引擎上获得了巨大成功。谷歌将这些经验汇聚成

基础。

GFS (2003)

件上,通过软件的方式自动容错,为大量的用户提供总体性能较高的服务。



图 4-2 Google 文件系统

缩性、可靠性以及可用性。但是，GFS 的设计还基于对应用的负载情况和技

术环境的不同，所以有了与传统文件系统明显不同的假设，并衍生出了完全不同的设计思路。例如，数据使用大到若干 GB 字节的大文件持续存储，而这些文件通常只是进行添加或读取操作。它也是针对谷歌的计算机集群进行的设计和优化，因为这些节点由廉价的“常用”计算机组成，必须防止单个节点的高损害率和随之带来的数据丢失。GFS 还采取了包括高数据吞吐率在内的其他设计理念，不惜造成访问反应期变差的后果。

架构上，一个 GFS 集群包含一个单独的 Master 节点、多台 Chunk 服务器，并且同时被多个客户端访问。Chunk 服务器存储数据文件，这些单个的文件像常见的文件系统簇或者扇区那样被分成固定大小的数据块。Master 节点通常不存储实际的大块数据，而是存储与大块数据相关的元数据。这样的数据如映射表格，将 64 位标签映射到大块数据位置及其组成的文件、大块数据副本位置、哪个进程正在读写特定的大数据块或者追踪复制大块数据的“快照”。所有这些元数据通过 Master 节点周期性地接收从每个数据块服务器来的更新以保持最新状态。

操作的允许授权是通过限时的倒计时“租期”系统来处理，Master 节点授权一个进程在特定的时间段内访问数据块。在这段时间内，Master 节点不会授权其他任何进程访问该数据块。当某个 Chunk 服务器被更改后（总是主要的数据块存储器），系统会将更改复制到其他 Chunk 服务器上。这些变化直到所有的 Chunk 服务器确认才存储起来，这样就保证了操作的完整性和原子性。

访问大数据块的程序首先会查询 Master 节点，得到所要数据块的位置。如果大数据块没有进行操作，Master 节点就回答访问程序大数据块的位置，然后程序就可以直接与 Chunk 服务器进行联系以接收数据。

谷歌的后面两篇论文——MapReduce 和 BigTable 都是以 GFS 为基础的，

三大基础核心技术构建出了完整的分布式运算架构。

MapReduce (2004)

第二篇论文是谷歌在 2004 年发表的《MapReduce: 超大机群上的简单数据处理》。

MapReduce 是一种编程模型, 如图 4-3 所示, 也是一个处理和生成超大数据集 (大于 1TB) 的算法模型的相关实现。其主要贡献是通过简单的接口来实现自动的并行化和大规模的分布式计算, 通过使用 MapReduce 模型接口实现在大量普通的 PC 机上高性能计算。MapReduce 使用 GFS 存储数据。

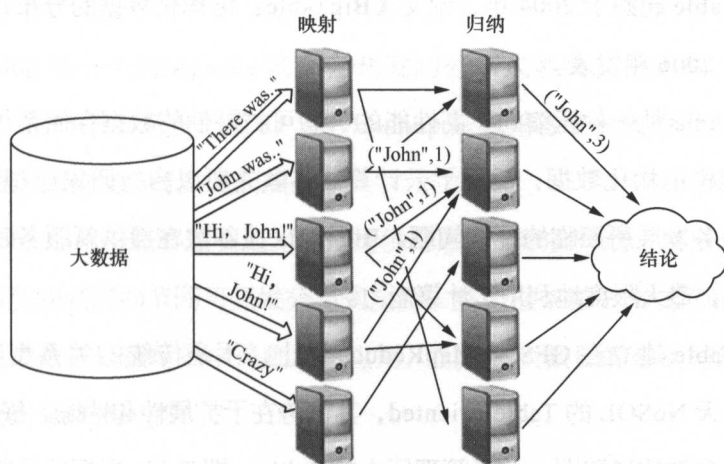


图 4-3 MapReduce 模型示意图

MapReduce 的主要思想——Map (映射) 和 Reduce (归纳) 来自于函数式编程语言 (如 Lisp)。其中, Map 处理的是杂乱无章的原始数据, 每条数据之间没有关系; 而在 Reduce 阶段, 数据是以 key 后面跟着若干个 value 来组织的, 这些 value 有相关性, 至少它们都在一个 key 下面, 于是就符合函

数式编程语言里 Map 和 Reduce 的基本思想了。

简单地说，面对由一些独立元素组成的概念，Map 操作就是对其列表上的每一个元素进行指定的操作。因为每个元素都是被独立操作的，而且原始列表没有被更改，所以 Map 是可以高度并行的。这对高性能要求的应用以及并行计算领域的需求非常有用。而 Reduce 操作是指对一个列表的元素进行适当地合并。因为 Reduce 总是有一个简单的答案，大规模的运算相对独立，所以 Reduce 在高度并行环境下也很有用。

BigTable (2006)

BigTable 起源于 2004 年，论文《BigTable：结构化数据的分布式存储系统》则于 2006 年发表。

BigTable 是一种压缩的、高性能的、高可扩展性的数据存储系统，用于存储大规模结构化数据，适用于云计算。谷歌之所以自行研发巨型数据库，是因为业务发展所面临的性能问题。BigTable 让谷歌在提供新服务时的运行成本降低，最大限度地利用了计算能力。

BigTable 建立在 GFS 和 MapReduce 之上，不是传统的关系型数据库，而更像今天 NoSQL 的 Table-oriented，其优势在于扩展性和性能。每个 Table 都是一个多维的稀疏图。为了管理巨大的 Table，把 Table 根据行分割，这些分割后的数据统称为 Tablets。每个 Tablets 大概有 100 ~ 200MB，每个机器约存储 100 个 Tablets。由于 GFS 是一种分布式的文件系统，采用 Tablets 的机制后可以获得很好的负载均衡。例如，可以把经常响应的表移动到其他空闲机器上，然后快速重建。

BigTable 的设计思路启发了无数的 NoSQL 数据库，如 Cassandra、

HBase 等。Cassandra 架构中有一半是模仿 BigTable，包括数据模型、SSTables 以及提前写日志。

谷歌的这三篇论文成了大数据开源化的主要推动力。

4.1.3 Hadoop 生态 10 年见证大数据时代

大数据技术起源于谷歌，谷歌发表了三篇关键论文，但并未直接将自己的软件开源。如果没有开源软件 Hadoop 的出现，就不会有如今大数据技术和应用的飞速发展。

Hadoop 简介

Hadoop 是一个由 Apache 基金会开发的分布式系统基础架构，诞生于 2006 年 1 月 28 日，至今已有 11 年。Hadoop 是根据谷歌发表的关于 GFS 和 MapReduce 的论文自行实现而成。

Hadoop 的框架能够透明地为应用提供可靠性和数据移动，用户无需了解分布式系统的底层细节即可开发分布式应用。整个框架在 ETL 方面的天然优势使其在大数据处理场景中得到广泛应用，加速了大数据的普及与发展，形成了广泛的 Hadoop 生态。

Hadoop 的演进

Hadoop 1.0 以及之前的早期版本主要由 HDFS 和 MapReduce 这两个核心组件（Core Hadoop）构成。HDFS 为海量的数据提供了存储，是对 GFS 的模仿。而 MapReduce 为海量的数据提供了计算，是对谷歌 MapReduce 的模仿。在 Hadoop 2.0 中，核心组件中增加了 YARN（Yet Another Resource

Negotiator)。YARN 是一个通用资源管理系统，可为上层应用提供统一的资源管理和调度。它的引入为集群在利用率、资源统一管理和数据共享等方面带来了巨大好处，把计算框架和资源管理彻底分开，改善了 Hadoop 1.0 在可扩展性、内存消耗、线程模型、可靠性和性能上的缺陷。

现在，Hadoop 已经从两个核心组件发展为包含 60 多个相关组件的庞大生态，其中包含在各大发行版中的组件就有 25 个以上。10 年来，这个生态不断成长，与 Linux 的生态非常相似。

Hadoop 发展至今，大致可以分为四个阶段。

➤ 诞生期（2002—2004 年）

缘起于谷歌的 GFS 和 MapReduce 两篇论文的发表，道格·卡汀（Doug Cutting）等人开发了开源网页爬虫项目 Nutch，并在其中实现了类似 GFS 的功能，即后来 HDFS 的前身。

➤ 孵化期（2005—2007 年）

Nutch 中接着又实现了 MapReduce 的功能。道格·卡汀加入雅虎，组建一个专门的团队，将 HDFS 和 MapReduce 移出了 Nutch 项目，并正式启动 Apache Hadoop 项目独立发展。之后，谷歌发表了 BigTable 论文，这也激发了 HBase 的创建。而随着第一个 Apache Hadoop 版本的发布，社区贡献者开始迅速增长。

➤ 扩散期（2008—2009 年）

Hadoop 的应用实践越来越广泛，升级成为 Apache 的顶级项目。不久，雅虎的 Hadoop 应用集群突破了 1 万个内核。之后，第一个 Hadoop 商业化公司 Cloudera 成立，并由道格·卡汀担任首席架构师。此外，子项目 Hive、Pig 也在这期间发布。

► 快速采用期（2010 年至今）

这一时期随着 Hadoop 生态的不断演进，项目不断发展。Spark 逐渐替代 MapReduce 成为了 Hadoop 的缺省执行引擎，它也和 HBase 一起共同成为 Apache 的顶级项目。另一个重要的企业级应用里程碑 YARN 成为 Hadoop 子项目。此外，第一个 Hadoop 原生 MPP 查询引擎 Impala，继 HBase 之后的第一个 Hadoop 原生存储替代方案 Kudu，相继加入 Hadoop 生态圈。

Hadoop 的未来

如今的 Hadoop 已经是一个生态和标准的代名词，如果要预测 Hadoop 的未来发展，我们就要思考哪些因素造就了 Hadoop 如今的成功。

首先是技术发展的整体需求。对计算要求的不断提升使以摩尔定律为代表的纵向扩展（Scale up）架构面临技术瓶颈，转而寻找以分布式为代表的横向扩展（Scale out）架构。廉价计算机集群具有很诱人的性价比，解决了技术上的难点后自然成为一个重要方向。

其次，从自身的技术框架选择来讲，“单一平台、多种应用”的理念以及“有弹性的多层级架构”是 Hadoop 领先其他产品的内因。Hadoop 在底层共用一份 HDFS 存储，上层有很多个组件分别服务多种应用场景。加上 Hadoop 生态的开源、开放和社区的多样、活跃，持续地推动了技术的发展，能够快速满足用户不断变化的需求，从而超越了同期的其他项目。

此外，以雅虎、谷歌为代表的工业界实践也给了 Hadoop 社区很大的信心。互联网公司的部署和实践已验证了 Hadoop 的可靠性和可用性，使其被更多的互联网公司和大型传统企业所采纳。大量的部署实践又促进了 Hadoop 的进一步发展和成熟，形成正向闭环。

那么，Hadoop 的未来又会如何？我们同样可以从这三方面来看。

首先，从整体技术发展上，现在处于云计算和物联网的时代，越来越多的数据将在云端进行计算，Hadoop 原来所推崇的本地化计算理念将面临挑战，而 HaaS (Hadoop as a Service) 将受到更多的关注和应用。物联网的发展则带来了海量的数据源，英特尔 CEO 科再奇 (Brian M. Krzanich) 曾经预测 2020 年将有 500 亿设备联网，会带来 50 万亿 GB 的数据；世界经济论坛预测 2022 年将有 1 万亿传感器入网，Hadoop 将不得不适应这种“数据规模爆炸”。

其次，成熟的架构拒绝通用型框架。随着各层的组件不断增多，目前的架构很难应对所有场景，未来的 Hadoop 可能会分裂为不同的数据量级和不同的应用场景，细分为更多的解决方案和产品组合。用户则根据自己的数据使用特性和计算需求选择相应的组合进行部署，如对于期望通过机器学习挖掘大数据价值的企业来说，拥有高级分析和实时处理等组件的解决方案将更受青睐。

最后，从工业界的实践上看，Hadoop 生态中的企业将更细分和聚焦。以 Cloudera、Hortonworks 为代表的企业是将 Hadoop 产品化，为大型企业提供 Hadoop 的使能服务。以 Acxiom、Tableau 为代表的企业则是为 Hadoop 生态创造更多有价值的工具，使整体生态溢价。以 Altiscale、Qubole 为代表的 HaaS 服务商面向小型公司提供云端化的 Hadoop 解决方案，丰富了 Hadoop 的用户覆盖面。此外，还有更多企业会将 Hadoop 应用于内部实践。

综上所述，至少在未来的 5 ~ 10 年，Hadoop 生态仍将处于蓬勃的发展期。

4.1.4 大数据技术发展趋势

Hadoop 生态是大数据的核心，但大数据技术远远不只是 Hadoop。

正如 Cloudera 的首席科学家、Hadoop 的创始人道格·卡汀所说：“我们在本世纪取得的大部分进展，是来自于对所产生的数据形成进一步的理解。”

将原始数据转化为分析洞察，一直是大数据技术最主要的推动力。数据量、数据种类和数据速度的提升，从各个方面带来了更强大的使用案例，智能分析、数据科学、数据安全、自助服务等将得到越来越多的关注。而人工智能、深度学习、VR、AR 和容器技术的发展，将使大数据的解决方案进入新的阶段，摩尔定律的边界将被不断延展。

4.2 区块链的技术沿革

4.2.1 区块链基础技术

点对点通信技术

点对点（P2P）通信技术是一种对等连接的互联网技术，依赖网络中参与者共同的计算能力和带宽，而不是把依赖都聚集在较少的几台服务器上。这类网络具有多种用途，各种档案分享软件已经广泛使用。点对点通信技术也被应用在类似 VoIP 等实时媒体业务的数据通信中。

纯点对点网络没有客户端或服务器的概念，只有平等的同级节点，同时对网络上的其他节点充当客户端和服务器。这种网络设计模型不同于传统的“客户端—服务器”模型，在“客户端—服务器”模型中通信通常来往于一个中央服务器，而中央服务器就是系统中的单点故障源头。点对点通信技术形

成的网络由于设计和实施方案的不同，也出现了不同的网络模型。

非对称加密算法

非对称加密算法需要两个密钥来进行加密和解密，其中一个公开密钥 (Public Key)，简称公钥，另一个是私有密钥 (Private Key)，简称私钥。公钥是可以公开的，也需要通过公共的渠道公开给相应的通信方；而私钥是非对称加密算法中的关键，需要安全地保护好。公钥与私钥是一对，如果用公钥对数据进行加密，只有用对应的私钥才能解密；如果使用私钥进行数据的签名，只有公钥可以来验签。

因为加密和解密使用的是两个不同的密钥，所以被称为非对称加密算法。由非对称加密算法形成的非对称加密体系解决了信息公开传送和密钥管理问题，是一种新的密钥交换协议，允许在不安全网络上的通信双方交换信息，安全地达成一致的密钥，这就是“公开密钥系统”。

区块 - 链式数据结构

区块链是一种把区块以链的方式组合在一起的数据结构，选取新节点时需要将新区块里前一个区块的哈希值、当前时间戳、一段时间内发生的有效交易及其梅克尔树根值等内容打包成一个区块向全网广播。密码学保证了数据的不可篡改和不可伪造，能够使参与者对全网交易记录的事件顺序和当前状态建立共识。

由于每一个区块的块头都包含了前一个区块的交易信息压缩值，这就使从创世块（第一个区块）到当前区块连接在一起形成了一条长链。如果不知道前一个区块的“交易缩影”值，就没有办法生成当前区块。因此，每个区块必定按时间顺序跟随在前一个区块之后。这种所有区块包含前一个区块的

引用结构，让现存的区块集合形成了一条数据长链。

分布式账本

在区块链网络中，每个新增加的区块里面存储了全网最近的交易记录，然后按时间顺序依次被添加到区块链上。区块链存储在由多个节点组成的分布式网络系统上，每个完整的节点都存储有整个区块链的副本，而每一个计算机节点之间通过网络进行交易信息的共享。

同时，区块链也是一个交易数据库，其中存储的是在系统中由所有节点共享的信息，称为分布式加密总账本。通过这个总账本，区块链实现了其不需要一个中央权力机构或受信任的第三方来协调互动、验证交易或监管行为的特征。一个区块链上的完整副本包含了每一个曾经执行的交易，使历史上的任何信息都可以被任何一个参加的节点所访问。全网共享账本这个特性也使防止双重支付成为现实。

105

共识机制

由于点对点通信的原因，分布式系统或者点对点网络中，每个节点是按自己的规则运行的。为保持整个系统数据的一致性，需要借助分布式共识技术。传统的分布式技术中，由于本身的技术特点并不需要数据的一致性，更多的是需要数据的准确性，包括主从、Sharding 技术等。但是，在点对点网络中每个节点之间并没有依赖关系，所以就更需要保证数据一致性的机制。

分布式共识算法包括异步和共识。异步是指在一个分布式系统中对消息的处理速度或者消息送达时间不做任何假设；共识是指当多个主机通过异步通信方式组成网络集群时，这个网络默认是不可靠的，那么在這些不可

靠的主机之间复制状态需要采取一种机制，以保证每个主机的状态最终达成相同一致性状态，即取得共识。通过共识机制，实现了区块链网络链条的唯一和数据的一致。

4.2.2 区块链更高层级的技术：资产互联

智能合约

在区块链的环境下，合约就是通过区块链使用密码货币和某人形成某种协议。传统意义上的合约，就是双方或者多方共同协议做或者不做某事来换取某些东西，合同中的每一方必须信任彼此会履行义务。智能合约的特点是，同样是彼此之间同意做或者不同意做某事，但无需再信任彼此，这是因为智能合约不但是由代码进行定义的，也是由代码（强制）执行的，自动完成且无法干预。

智能合约之所以能如此操作，主要是由三个要素造成的：自治、自足和去中心化。自治表示合约一旦被启动就自动执行，而不需要它的发起者进行任何干预；智能合约能够自足地获取资源，也就是说通过提供服务或者发行资产来获取资金，当需要的时候也会使用这些资金；智能合约是去中心化的，它们并不依赖单个中心化的服务器而分布式存在，并且通过网络节点来自动运行。

虚拟机

虚拟机是区块链中智能合约的运行环境，它被沙箱封装起来，完全隔离。也就是说，运行在虚拟机内部的代码不能接触到网络、文件系统或者其他进程，甚至智能合约之间也只能进行有限的调用。

4.3 大数据和区块链技术之间

4.3.1 分布式数据库：HDFS vs. 区块

大数据需要应对海量化和快速增长的存储，这要求底层硬件架构和文件系统性价比上要大大高于传统技术，能够弹性扩张存储容量。谷歌的 GFS 和 Hadoop 的 HDFS 奠定了大数据存储技术的基础。另外，大数据对存储技术提出的另一个挑战是多种数据格式的适应能力。因此，现在大数据底层的存储层不只是 HDFS，还有 HBase 和 Kudu 等存储架构。

区块链本质上是一种分布式的数据库系统。区块链技术作为一种链式存取数据技术，通过网络中多个参与计算的节点来共同参与数据的计算和记录，并且互相验证其信息的有效性。从这一点来说，区块链技术也是一种特定的数据库技术。由于去中心化数据库在安全、便捷方面的特性，很多业内人士看好区块链的发展，认为区块链是对现有互联网技术的升级与补充。

4.3.2 分布式计算：MapReduce vs. 共识机制

大数据的分析挖掘是数据密集型计算，需要巨大的分布式计算能力。节点管理、任务调度、容错和高可靠性是关键技术。谷歌和 Hadoop 的 MapReduce 是这种分布式计算技术的代表，通过添加服务器节点可线性扩展系统的总处理能力，在成本和可扩展性上都有巨大的优势。现在，除了批计算，大数据还包括流计算、图计算、实时计算、交互查询等计算框架。

区块链的共识机制，就是所有分布式节点之间怎么达成共识，通过算法来生成和更新数据。认定一个记录的有效性，既是认定的手段，也是防止篡改的手段。区块链主要包括四种不同的共识机制，适用于不同的应用场景，在效率和安全性之间取得平衡。以比特币为例，采用的是工作量证明，只有在控制了全网超过 51% 的记账节点的情况下，才有可能伪造出一条不存在的记录。

4.3.3 分布式和集中式技术的螺旋发展

和人类社会一样，IT 技术的发展也呈现出“合久必分，分久必合”，即集中与分布的螺旋式上升。

计算机诞生初期，仅能实现一对一的使用，是集中化的。为了使一台大型机能够同时为多个客户提供服务，IBM 公司引入了虚拟化的设计思想，使多个客户在同时使用同一台大型机时，就好像将其分割成了多个小型化的虚拟主机，是时分复用的集中式计算。

进入小型机和 PC 时代，回归了一对一的使用，不过设备已经分散到了千家万户。进入互联网时代，C/S 模型的客户端和服务端是分布式计算，只不过服务器之间还是分散的。

进入云计算时代，计算能力又被统一管控起来，在客户端和服务器的分布式计算基础之上，服务器之间也开始了分布式协同工作。因为协同，所以也可以认为它们在整体上是一种集中式的计算服务。

进入大数据时代，云计算成为大数据基础设施，也使大数据的核心思想和云计算一脉相承。MapReduce 将任务分解进行分布式计算，然后将结果合并，从而实现了信息的整合分析。

区块链是纯粹意义上的分布式系统，而分布正是技术的信仰。技术的背

后，还有商业和历史。

商业信仰集中，希望通过产品实现更好地控制和更高的利润。但随着产品集中度的不断上升，系统会越来越复杂，实现的难度越来越大，沟通、交流和管理成本也越来越高，最终变得不经济。

而历史告诉我们，人类社会存在社会化分工，让专业的人做专业的事。涂尔干的《社会分工论》谈到，“分工使社会像有机体一样，每个成员都为社会整体服务，同时又不能脱离整体。分工就像社会的纽带，故谓之‘有机团结’。”

基于这种思想，分布式技术诞生，产品功能被分解并分布到不同的节点上去完成，节点之间通过网络实现沟通。分布式系统中的一些节点或因为商业上的成功，重新成为“集中化”的节点。但随着时代的改变，它们终将会进入新一轮的分布式周期，如此往复。

集中和分布不是光谱的两端，任何伟大的产品都是商业和技术的“有机团结”。

109

4.4 大数据和区块链技术之异

4.4.1 两个技术处于不同的生命周期

技术成熟度曲线是咨询公司高德纳用来分析和预测各种新科技成熟演变速度及所需时间的著名工具。大数据与区块链在高德纳历年技术成熟度曲线中的出现情况如图 4-4 所示。

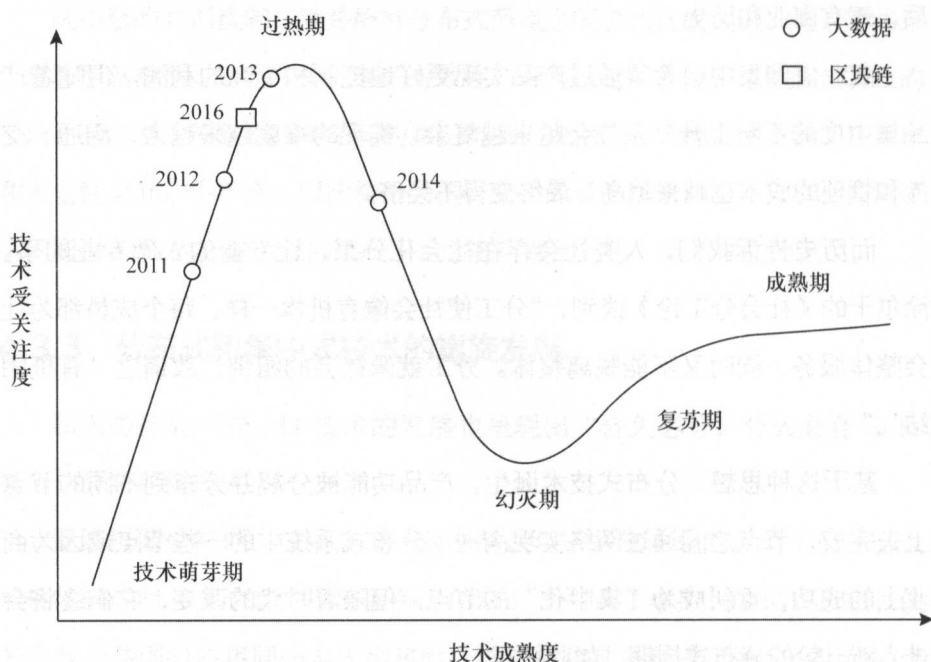


图 4-4 大数据与区块链在高德纳历年技术成熟度曲线中的出现情况

大数据于 2011 年第一次上榜，位于技术萌芽期的爬坡阶段，当时还统称为“‘大数据’和极端信息处理和管理”（“Big Data” and Extreme Information Processing and Management）；2012 年更进一步，并在 2013 年几乎达到了过热期顶峰；经历了 2014 年的下滑，从 2015 年开始，大数据突然从曲线中消失，可解读为高德纳对大数据的定位已从“新兴”转为“主流”。当前，大数据对于企业的意义已从能力要素上升为战略核心。

相对而言，区块链直到 2016 年才第一次出现在技术成熟度曲线中，并直接进入过热期。从总体来看，大数据和区块链所处生命周期的阶段大不相同，两者约有 5 年左右的差距。

4.4.2 大数据和区块链的主要差异

大数据通常用来描述数据集足够大、足够复杂，以致很难用传统的方式来处理。目前区块链能承载的信息数据是有限的，离大数据标准还差得很远。区块链与大数据有以下几个显著差异。

► 结构化 vs. 非结构化

区块链是结构定义严谨的块，通过指针组成链，是典型的结构化数据，而大数据需要处理的更多是非结构化数据。

► 独立 vs. 整合

区块链系统为保证安全性，所承载的信息是相对独立的，而大数据着重的是信息的整合分析。

► 直接 vs. 间接

区块链系统本身就是一个数据库，而大数据是对数据的深度分析和挖掘，是一种间接的数据。

► 数学 vs. 数据

区块链试图用数学说话，主张“代码即法律”，而大数据试图用数据说话。

► 匿名 vs. 个性

区块链是匿名的，而大数据强调的是个性化。

4.4.3 差异能否调和

对一个分布式系统来说，存在 CAP 定理（CAP theorem），又被称作布鲁尔定理（Brewer's theorem），它指出一个分布式系统不可能同时满足以下三点。

➤ 一致性 (Consistence)

分布式系统中的所有数据备份在同一时刻是否有同样的值。

➤ 可用性 (Availability)

集群中的一部分节点发生故障后，集群整体是否还能响应客户端的读写请求。

➤ 分区容忍性 (Partition tolerance)

当集群中的某些节点无法联系时，集群整体是否还能继续进行服务。

由于当前的网络硬件仍然会出现延迟丢包等问题，所以分区容忍性是必须要实现的。换句话说，CAP 定理表明必须在一致性 (C) 和可用性 (A) 之间进行权衡。

具体到区块链和大数据来说，大数据是以牺牲一致性 (C) 来换取可用性 (A) 和分区容忍性 (P)，而区块链却优先保证了一致性 (C)。

4.4.4 可相互借鉴之处

通过 CAP 定理，我们知道区块链和大数据的诸多特性无法两全，需要针对具体场景，在多样化的取舍方案下设计出多样化的系统。

区块链 + 大数据：在区块链中使用大数据技术

区块链是一种不可篡改的、全历史记录分布式数据库存储技术，巨大的区块链数据集合包含了每一笔交易的全部历史。随着区块链技术的迅速发展，数据规模会越来越大，不同业务场景的区块链数据融合会进一步扩大数据规模和丰富性。

区块链以其可信性、安全性和不可篡改性让更多数据被解放出来，推

进了数据的海量增长。区块链的可追溯性使数据的质量获得前所未有的强信任背书。通过区块链脱敏的数据交易流通，则有利于突破信息孤岛，并逐步形成全球化的数据交易。

区块链提供的是账本的完整性，数据统计分析的能力较弱。大数据则具备海量数据存储技术和灵活高效的分析技术，极大地提升了区块链数据的价值和使用空间。

大数据 + 区块链：在大数据中使用区块链技术

大数据的技术生态百花齐放，没有哪个软件能解决所有的问题，能解决问题也是在一定范围内，即使是 Spark、Flink 等知名软件。在强调透明性、安全性的场景下，区块链有其用武之地。在大数据的系统上使用区块链技术，可以使数据不能被随意添加、修改和删除。当然，其时间和数据量级是有限度的。

113

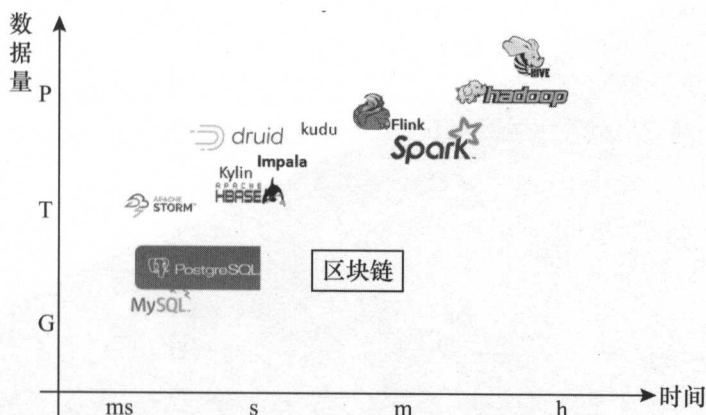


图 4-5 区块链在数据处理软件 / 平台中的位置

图 4-5 以时间、数据量为坐标轴，列出了目前大数据引擎大致擅长处理

数据的范围，区块链可在其中成为一种很好的补充。

例如，存档的历史数据因为是不能被修改的，所以可以对大数据作哈希处理并加上时间戳，存在区块链上。未来当我们需要验证原始数据的真实性时，可以对相应的数据做同样的哈希处理：如果得出的答案相同，则说明数据是没有被篡改过的。也可以只对汇总数据和结果做处理，这样只需要处理增量数据，那么应对的数据量级和吞吐量级可能是今天的区块链或改善过的区块链系统可以处理的。

随着数字经济时代的大数据能够处理越来越多的现实预测任务，区块链技术能够帮助把这些预测落实为行动。通过把区块链技术与大数据相连接，大数据将会在“反应—预测”模式的基础上更进一步，能够通过智能合约和未来的 DAO、DAC 及 DAS 自动运行大量的任务，那时将会解放大量的人类生产力，让这些生产力被去中心化的全球分布式计算系统代替。

清华大学出版社

北京清华大学出版社

第5章

互联互通，引领变革

数

据作为重要的资源，受到越来越多的关注。
在数据资源开发利用的过程中，分享总是难题：政府束手束脚，担心泄露机密；个人也有所顾虑，担心自己的隐私被曝光；企业更是把院门关得紧紧的。

但就像现代贸易极大地促进了全球的生产和消费一样，WTO推动了中国经济社会的高速发展，数据也需要流动起来才能体现价值，发挥作用。没有流通的数据，作用很小，甚至只是一堆数字。

就像经济活动的驱动力是价值实现，作为同样有价值的数据，本身在流动过程中就需要有对等的价值流动。而这正是传统的大数据行业容易忽视的地方，也让大数据行业一直得不到很好的发展。

带着数字密码货币基因的区块链本就是为价值所生，有能力补上大数据价值流转这块短板。只有全面实现价值的大数据，才会真正拥有未来！

5.1 数据流通

5.1.1 大数据经济来临

数据社会的基础是数据经济

大数据时代各行各业都主动运用数据思维进行转型和创新，整个社会的生产模式、交付方式、生活体验和管理决策能力都在向“数据社会化”演进。数据能够平等地被社会各层面使用，打破物理疆界，渗透到社会生活的每个角落，进而驱动虚拟世界与现实社会之间实现生态交互，让社会资源能够在同一平台上被重新整合、共享和使用，最终实现全部的社会应用价值。

因此，未来的新经济模式将以信息经济、知识经济、智慧经济为核心，这些都需要以社会化的大数据为基础，以社会数据的充分融合升华为条件，以大数据资源的商品化运行和有偿使用为标志。只有真正让大数据成为有效的资源，被经济规律所引导且又能服务于经济社会，才能真正实现大数据更高层面的充分使用，并真正奠定数据社会的基础。

数据经济的核心是数据流通

大数据最吸引人的地方在于“数据外部性”，即同一组数据可以在不同的维度上产生不同的价值和效用，对不同的用户也会发挥不同的效用。因此，

使用维度增加,数据的能量和价值就将层层放大。同时,数据可以边际成本很低地被“复制”,本质上是天然过剩的,所以可以跨越时间和空间进行更大程度的重复利用,从而形成更大的社会效用。

互联网时代,最著名的网络效应评估方法是梅特卡夫定律¹。该定律认为一个互联网的价值与其节点数的平方成正比,其中有个关键假设是这些节点间的关系具有同等价值。这个定律也成为了互联网经济的基石。

那么,数据经济的基石又是什么?数据的流通可以带来很大的价值,这里的数据流通不仅包括数据的交易和交换,同时也包括数据的开放和共享。数据的顺畅流通将有效降低创新门槛,带动移动互联网、大数据及数据服务等新兴产业的发展,成为数据经济的引擎。在大数据上升为国家战略的背景下,数据流通的发展更是拥有了市场和政策的双重机遇。

5.1.2 数据流通的发展与预期差距较大

数据流通的引爆点尚未到来

前文已经对数据开放方面遇到的诸多困难进行了分析,此处不再赘述。从数据交易方面来看,数据流通还有很大的提升空间。全国已有 20 多家各类数据交易平台,如表 5-1 所示,但真正的交易量却少之又少。即使目前走在业界前沿的贵阳大数据交易所高达几千万的交易额,但这些交易大部分都是点对点的线下撮合交易,这样的交易额远远没有达到大家预期中的大数据广泛交易的盛况。

1 梅特卡夫定律是由 3Com 公司创始人、“以太网之父”梅特卡夫提出的。

表 5-1 目前中国主要大数据交易平台列表

交易所名称	地点
贵阳大数据交易所	贵阳
上海数据交易中心	上海
长江大数据交易中心	武汉
浙江大数据交易中心	杭州
哈尔滨数据交易中心	哈尔滨
华中大数据交易平台	武汉
钱塘大数据交易中心	杭州
武汉东湖大数据交易中心	武汉
华东江苏大数据交易平台	盐城
安徽淮南数海大数据交易平台	淮南
阿里云数据市场	杭州
百度 APIStore	北京
京东万象	北京
数据堂	北京
优易数据	北京
类聚	济南
数读	广州
发源地	上海
数据星河	贵阳
数据宝	贵阳
数据淘	成都
大海洋数据服务平台	合肥
北京大数据交易支撑平台	北京
大数据采集工具	合肥

大数据价值的发挥在于多源数据的融合，并根据不同的应用需求做出数据产品。目前的数据流通市场仍未引爆，严重制约了社会整体大数据价值的发挥。

数据流通亟待解决的几个痛点

数据流通非常难，第1章已经详细分析了当前阻碍数据共享乃至影响大数据发展的近忧与远虑，本节将从实际操作的角度对几个具体问题进行分析。

► 数据权属：无法清晰界定

如何定义数据的权属并不是一件容易的事，涉及技术、商业和法律等多方面。在产权不清晰的前提下，拥有数据的主体没有动力将数据分享出去，否则会带来自身利益的损耗。如果无法保护数据产权，数据一旦出售就会面临被无限次倒卖的风险，数据的市场价值也因无限的供给量而骤减。

当前技术条件下，无法清晰界定数据的所有权和控制权。行业潜规则是“谁采集，谁拥有”，企业将客户在其网站和 App 等载体上所生产的数据当成自己的资源，而生产数据者却无法有效控制自己生产的数据。用户每天在各种交易、社交网站和 App 上产生大量的数据，这些都是用户未来重要的信用资源，现在却完全无法为用户本人所控制。

因此，许多学者认为要加快数据领域的制度建设，并提出了数据应赋予5个基本权利，如表5-2所示。

表 5-2 数据权利列表

数据权利	说明
数据拥有权	必须像其他物理财产一样明确数据的拥有权，拥有权可以变更和分割
数据隐私权	明确什么数据能够披露、什么数据不能披露、披露到什么样的粒度
数据许可权	哪些人在什么时间有权利看数据是有约束的。例如，今天允许给某个人看数据，明天就不允许。这个权利是可撤销的，也是可转移的
数据审计权	需要有一种审计机制来监督用户按照某个规范许可使用数据，确保用户按照约定的许可规范使用数据
数据分红权	基于数据外部性，获得数据使用许可的一方在反复使用数据时会产生新的价值，数据拥有者有权利要求获得回馈

➤ 数据质量：标准不一，良莠不齐

从小数据时代开始，不同来源的数据就各有各的格式。而在大数据时代，由于数据源的千差万别，采集的数据无论格式还是质量都存在很大的差别。一方面，即使相同格式的数据，也可能存在语意和度量衡的差别，如同形状不一的石块很难直接垒成摩天大楼。另一方面，原始数据会有缺漏和错误之处，也可能混有大量无效和垃圾数据，必须使用一些手段进行数据的清洗，否则无法使用。

当有一天数据量足够大时，也许这两个问题会随之消失。

➤ 数据安全：隐私和滥用无法保障

数据安全问题是保障数据权属的核心。数据未经生产主体同意而被采集并使用，甚至流入数据黑市，会造成用户安全、企业安全甚至国家安全方面的连锁反应。但现在数据被私自采集和滥用的现象十分普遍，这反过来导致很多数据主体参与数据流通的意愿不强。

在数据采集、分析技术进步的同时，整个社会也变得愈发透明，“被监视”的范围被无限扩大。

➤ 数据定价：数据价值无法准确衡量

有了数据权利和保障数据安全这两大前提，数据才能进行定价。

数据已经被广泛认可是一种资产，具有无形财产和资产的属性。但数据的价值应如何准确衡量呢？目前，数据资产进入资产负债表还不太可行。对数据价值的财务量化，已有机构提出需要从数据的内在价值、业务价值、绩效价值、成本价值、市场价值以及经济价值等维度考虑，涉及数量、范围、质量、粒度、关联性、时效、来源、稀缺性、行业性质、权益性质、交易性质、预期效益等因素。通过衡量各个因素的权重配比、不同的指标量级，才能实

现对数据资产的全方位、标准化评估。

目前，数据定价尚无成熟的方法，定价时主要的依据有两个：一是根据效用，即数据使用的频率，从分析结果逆推数据的渊源，从而量化各方数据对结果的贡献度；二是根据稀缺性，即根据数据价值的密度以及历史价格的稀缺性进行定价。还有学者提出了应用博弈论、人工智能等方法对数据资产进行评估的观点，但是都不能很好地解决数据价值量化的问题。

从更广泛的角度看，对数据的估值应该考虑估值七律。

(1) 数据可以被无限次共享而不损失价值，但多次复制会使所有权复杂化，增加成本。

(2) 数据用得越多，价值越大，不像许多资产（如车、厂房和设备）在使用中会贬值。

(3) 数据价值会随时间衰变，不像酒，保存得越久，价值越高。

(4) 数据越精确，价值越大。

(5) 多个独立数据源的融合会带来 $1+1>2$ 的效果。

(6) 更多的数据不一定能带来更多的价值。

(7) 数据不会损耗，反而越用越多。相反，数据如果不使用，就会变成负债。

这些基本原则对数据的定价具有指导意义，对数据的使用频率、新鲜度、质量、外部性都是重要变量，特别是在移动互联网时代针对具有高频性、时效性、随机性等鲜明特点的各种数据。

5.1.3 区块链能带来哪些改变

区块链技术凭借不可篡改、可追溯等特性，可以解决数据共享开放与交

易交换中的若干关键问题。

（1）数据权属：区块链可以提供可追溯路径，能有效破解数据确权难题。

区块链对数据进行注册、认证，确认了大数据资产的来源、所有权、使用权和流通过程，让交易记录透明、可追溯和被全网认可。把各个区块的交易信息串起来，就形成了完整的交易明细清单，每笔交易的来龙去脉非常清晰、透明。区块链使数据作为资产进行流通时更有保障，有助于让数据真正实现资产化。

简单地说，数据一旦上链，便永远带有原作者的“烙印”。即使在网络中经过无数次复制、转载和传播，还能明确数据的生产者和拥有者。数据的接收者对数据本身或交易情况有任何疑问，还可以根据记录进行查询和追溯。区块链能够进一步规范数据的使用，精细化授权范围，把数据的所有权还给用户自己。

（2）数据质量：制定数据标准，并通过共识验证改善数据质量。

区块链对数据进行注册和认证时有明确的格式要求，从而能够明确该链数据的语意和度量衡，一方面能够统一单条链的数据标准，另一方面在多源数据进行融合时能够实现快速清晰的解读。

同时，区块链的数据溯源机制可以改善数据的可信度，让数据获得信誉。多方可检查同一数据源，甚至通过给予评价来表明他们认为的数据有效性。区块链使数据的质量获得前所未有的强信任背书，也保证了数据分析结果的正确性和数据挖掘的效果。区块链共识验证数据，也是梅兰妮·斯万²所提到的最高推荐等级的数据，因为这个精度和质量是基于群体共识来支持的。

（3）数据安全：以多种加密技术保障数据安全和隐私。

数据安全是数据互联的基础，有多方面的含义：一是数据本身的安全，

2 梅兰妮·斯万：《区块链：新经济蓝图及导读》的作者。

主要是指采用现代密码算法对数据进行主动保护，如数据保密、数据完整性、双向强身份认证；二是数据防护的安全，主要是采用现代信息存储手段对数据进行主动防护，如通过磁盘阵列、数据备份、异地容灾等手段保证数据的安全；三是数据的访问控制，包括各种权限控制，以保证数据的访问是在授权范围之内；四是数据处理的安全，是指如何有效地防止数据在录入、处理、统计或打印中，由于硬件故障、断电、死机、人为的误操作、程序缺陷、病毒或者黑客等造成的数据损坏或数据丢失现象，以及某些敏感或保密的数据被没有资格的人员查看而造成数据泄密等后果。

当数据被哈希处理后放置在区块链上，数字签名技术使只有获得授权的人才可以对数据进行访问；通过私钥既能够保证数据私密性，又可以共享给授权研究机构；数据统一存储在去中心化的区块链上，在不访问原始数据的情况下进行数据分析，既可以对数据的私密性进行保护，又可以安全地提供给研究机构 and 研究人员共享。

系统安全和数据安全还需要审计监控作为保证。通过区块链的智能合约，可以给出数据使用的具体条款，并照此监督数据的使用。条款必须有形式化的描述，其目的在于让非 IT 专业人员能够编写这些条款，如企业法务。如果企业的数据要流通，需要法律人士给出逻辑严密的使用条例，条例的内容本质上不属于 IT 范畴。对于个人用户，通过审计监控和精细化授权也能最大程度地保护用户隐私。

在企业内，使用区块链技术合并来自不同区域办公室的数据，不但能降低企业审核自身数据的成本，还可以与审计员共享数据。在某个生态系统内，如银行，竞争对手过去永远不会分享他们的数据，但现在可能会坦率地展示。因为结合几个银行的数据就可以做更好的模型以预防信用卡欺诈，或者供应

链机构通过区块链共享数据可以更好地支持供应链运转。在全球范围内，区块链可以促进不同生态系统之间的数据共享。在某些情况下，当孤立的数据被合并，不只可以得到一个更好的数据集，还可以得到一个新的数据集，从中可以收集到新的见解、新的业务应用。也就是说，以前做不到的事情现在也许可以做到了。

（4）数据定价：明确交易历史和各方贡献，助力数据价值衡量。

未来的数据市场需要有灵活的数据定价模型，既考虑数据的使用历史和时间变化所形成的基础价值，又能计量当前这次使用中可量化的价值，计算出这次交易的数据定价。同时，如果这次使用的是多方数据，可以根据各方贡献的大小对其数据进行分别定价。

区块链的可追溯性和不可篡改性能够明确数据的使用历史和交易历史，有助于衡量各方的贡献，从而设计出更灵活的数据定价模型。例如，将一次定价变为多次定价，根据一定时期内数据所发挥的价值，按周期对各方的贡献进行“分红”。

（5）数据支付：对数据的使用和流通进行快速、便捷的即付即用。

由于数据的多源性和复杂性，采用一个传统的交易平台无法实现对具体数据的精准流通，只能采用打包的方式，一方面影响了数据的质量，另一方面也降低了数据的价值。

因此，区块链的出现会从两个方面产生巨大的变革。第一，由于区块链本身带价值传输也就是支付的功能，基于区块链的交易系统可以很方便地实现数据的交换和支付。第二，区块链交易具有廉价、高频的特点，可以实现数据交易的细分。用户可以对所需要的数据实现精准支付，以满足自身对数据质量的要求。而数据提供者也可以采用更精确的定价，以使整个数据链

上的每一方都受益。

5.1.4 打造可信任的数据资产流通环境

在中心化的数据流通模式中，中心节点有条件、有能力复制和保存所有流经的数据，这对数据生产者极不公平。这种隐患仅仅依靠承诺是无法消除的，也成为阻碍数据流通的巨大障碍。基于去中心化的区块链能够破除中心节点拷贝数据的潜在威胁，有利于建立可信任的数据资产流通环境。

这里有必要参考 Windhover 公司提出的关于数字身份、信任和数据的茶隼原则（Windhover Principles）。这是一个由公共和私营部门的利益相关者合作编写的原则性框架，其目的是保护个人身份、信任和对互联网共享公开数据的获取。

原则 1：对个人身份的权利和个人数据的控制。

原则 2：透明地实施与有效地治理。

原则 3：确保信任与隐私。

原则 4：开源协作。

原则 1 意在让个人控制自己的数字身份凭证和个人资料，而不是由社会网络、政府或企业来掌控。构建数字社会，加强创新数字技术，应该以提高隐私的管理和实施为首要任务。例如，各人的身份证信息被政府、银行、医院等机构共享，如果某个机构泄露了，这些信息就为全世界所共享，那么这违反了原则 1。

原则 2 意在加强对个人隐私的管理，改善可调节性的法律审计并加强法律的执行力度。为了满足原则 1，任何组织只能知道个人的一个“短期”或“部分”身份信息，而无法拥有“永久”或“全部”身份信息。例如，医院可以

知道健康记录，但不能知道财务记录。在大数据时代，各个机构都想要关联一个人的多种信息。因此，我们需要原则2。

原则3意在建立一个有效的自主识别系统，需要加深隐私、信任、安全、治理、问责等保护机制。当有动机不纯者不断尝试进入系统时，任何安全和保护机制必须不断改进。所以，我们又需要原则3。

原则4意在用一个包容、开源的方法构建系统来体现这些规则，让人们信任这样的安全机制。区块链系统有开源的，也有不开源的。但区块链技术本身的运行机制保证了只有数据权利人才真正拥有数据，达到了可信任系统的标准。

茶隼原则的最终目的在于，个人的身份凭证能够为自己所有，由自己掌控。而如果将每个数据个体视为分布式自治组织，这个原则也将有利于建立可信任的数据资产流通环境。

在未来高度分布、去中心化场景下，区块链将承担全球去中心化数据流通中的数据资产管理、交易、支付、智能合约等业务，区块链技术也将被应用于个人数据控制（如 Open Mustard Seed 框架）和分布式数据存储（如 MaidSafe）。

未来的互联网数据层很有可能构建在区块链的基础上，并在上面叠加网络层、应用层，成为下一代互联网基础设施平台。未来的数据流通也许就构建在这些新的基础设施之上。

5.1.5 展望

大数据行业发展到今天，取得了令人振奋的成果，但也面临着数据流通相对不足的巨大挑战。区块链以其可追溯性、安全性和不可篡改性，在数据

流通的若干痛点上都能起到明显的改善作用，有利于降低信息摩擦、突破信息孤岛，让更多数据被解放出来。

区块链保障并促进了数据的流通，而流通带来了大数据困局的破解，将逐步推动形成社会化的数据流通网络。同时，区块链还能促进更平等和自由的数据流动，所产生的基于共识的数据具有更致密的价值属性。因此，区块链很可能将成为一个提升人类社会信息精度的工具。

5.2 价值互联

没有明确价值属性的大数据只是一堆数据而已。本质上讲，现阶段的大数据发展中遇到一些瓶颈和挑战，和大数据自身的价值（特别是对不同维度）不能得到很好的体现是有关系的。利益关系的不明确，以及利益实现渠道被阻断，导致大数据相关主体很难形成合力，突破阻碍来推动大数据的发展。

区块链技术由于在权属确定、数据安全、交易灵活性方面的优势，因而对数据价值的展示和实现有明显的帮助。区块链和大数据的结合，可以为未来社会的价值互联做出巨大的贡献。

本节具体谈一下区块链在数据资产的价值互联方面的诸多关键性优势。

5.2.1 以比特币为代表的去中心化性

在中心化的体系中，授权机制是一种信任机制，即默认管理人都是尽责和完善的。然而，比特币采用的是彻底非信任机制，其中每一个节点都是平等的，在足够多的节点确认账本交易之后便产生了比特币系统自身的信用价

值，从而解决了“双花”这样的问题。比特币的非信任机制正是去中心化思想的体现。

在非信任机制中，博弈是非常重要且必要的。比特币通过“挖矿”POW 机制来激励“矿工”维护系统和记录交易，通过开源社区的方式来监督并规范开发者的开发更新，通过自愿交易手续费的方式让用户更有效地使用比特币平台。

“去中心化”容易被大家误解为就是节点的分散、数据的分散、“矿工”的分散、开发者的分散……而比特币创始人中本聪是反对“一 IP 一票制”的，因为每位“矿工”的电脑都贡献为一个全节点，相当于网络节点的所有 IP 都拥有相等的权力。那么，那些拥有分配大量 IP 地址权力的人，如僵尸网络就可能主宰比特币网络。因此，比特币最后出现的是“一 CPU 一票”，实际是说一个计算单位代表着一个权力单位，拥有的计算力更高即意味着更高的权力。这是 POW 机制中“计算即权力”思想的形象化表达。

129

“去中心化”实际是对过程的描述，而不是状态，可以理解为是一种程序公正，而不是结果的绝对公平。所以，“去中心化”讲的是每个人参与共识的自由度问题，在代码开源、信息对称的前提下参与和决策的自由度，即意味着公平。

“去中心化”并不是什么新词，它其实就是亚当·斯密的那只看不见的手——市场的自由竞争。在竞争机制下，算力的集中并不是什么可怕的问题。一方面，由于高昂的计算力成本，“矿池”“矿工”发起 51% 攻击不符合理性经济人的前提；另一方面，即使存在不可理喻的疯子，如拥有大量算力份额的“矿池”，他们的攻击也不可持续，因为“矿池”的算力并不真正属于他们自己，而且随时面临新加入的算力、新玩家的挑战。算力集中本身就是市场

的结果，任何一个开放系统在自由竞争下都会形成专业化分工，这就好比生物有机体的组织分化。专业化的“矿工”、专业化的支付钱包、专业化的区块链数据服务商等，正是区块链去中心化的结果，而不是我们处心积虑要避免的后果。

建立不依赖第三方信任、不可操纵的去中心化的交易机制，是区块链在价值互联体系里的第一个特点。

5.2.2 匿名性

区块链系统中，节点之间的交换遵循固定的算法而无需信任。因此，交易对手无需公开身份来获取信任，这对节点的信用累积是非常有帮助的。

在计算机科学中，匿名是指具备无关联性的化名。所谓无关联性，就是指站在攻击者的角度，无法将用户与系统之间的任意两次交互进行关联。在比特币中，由于用户反复使用公钥哈希值作为交易标识，交易之间显然能建立关联。因此，比特币并不具备匿名性。

区块链技术的匿名性，是因为不知道每个地址对应的实际持有人是谁。你能证明这个地址是你的，方法就是对这个地址进行签名。只有在持有比特币地址私钥的情况下，才能对指定的地址进行签名。

利用区块链自身的技术优势，保证整个区块链体系的匿名性，脱离身份背书来参与交易，并在网络上积累不可篡改的信用记录。这是区块链在价值互联体系里的第二个特点。

5.2.3 支付便捷性

随着支付宝、微信支付等移动端支付方式的兴起，出门只带手机而不带

钱包已成为一种常态。但放眼全球，跨境汇款还远不能如此轻松潇洒，人们依然面临着费用高昂、有时间延迟、容易出错等问题。

据世界银行统计，全球跨境支付规模以年均 5% 的速度增长，2016 年已达到 6010 亿美元，中国有望超越巴西成为继美国和欧元区之后的第三大支付地区。然而，目前跨境汇款的成本高昂，平均每个汇款人所承担的手续费率达 7.68%。

Swift 系统的汇款流程如下。

第一步，机构在接到汇款需求之后要先进行反洗钱核查以及身份核验，收取资金以及相应费用，再开始处理汇款人的请求。

第二步，在转移环节，由于不是所有银行都加入了 Swift 联盟，银行或者汇款机构只得采取两种渠道进行跨境资金转移：Swift 会员可以利用 Swift 网络进行，而非会员则只能通过当地的代理银行进行汇款。

第三步，收款人收到通知后去银行或者汇款机构。

第四步，银行或者汇款机构进行身份核验。

第五步，收款人以当地货币获得汇款。

最后一步，根据当地监管政策，银行或者汇款机构有时需要向监管层提交包括交易细节在内的报告。

其实，整个 Swift 体系是在 20 世纪 70 年代建立的，在当年的技术背景下还算高效快捷，但在互联网信息革命后就难免落伍了。

比起传统的交易体系，区块链技术可以让交易更迅速、成本更低。以区块链技术为基础的全球网络可以接近实时地实现汇款的划转，并仍然保持本来区块链具有的匿名、可追溯、不可篡改等优势。而区块链网络交易环节单一、无需其他代理中介的参与，让全球汇款成本趋近于区块链网络的普通交易成

本。无论是交易速度、交易安全，还是交易成本，区块链系统颠覆性地改变了由来已久的全球汇款模式。实际上，2016 年底和 2017 年初，比特币价格大涨创出历史新高，就有传言是成为了全球汇款的一个通道来回避中国的外汇管制。

互联网化的交易流程和交易撮合，让区块链业务模式和传统的汇款等银行业务模式出现了鲜明的对比，体现了巨大的优势。这是区块链在价值互联体系里的第三个特点。

5.2.4 不可篡改性

区块链是一种“共识”实现技术，通过区块链可以记录全网所有的交易，供区块链的用户见证实现“共识”，而且链上信息内容不可篡改。这种不可篡改性是通过系统内多个副本的存在增加了内容被恶意篡改的成本。

当所有记录得到公示，就解决了现实生活中的“两表不可测”问题。两表不可测的原因是没有中心，两表数值各异，观者无从可知。区块链上存在多个节点，每个节点都有一个表。当存在多个表，而且多数指针指向一个时间，少数服从多数，观者也就心知肚明。

所以，区块链要解决的一个问题就是“少数服从多数”，少数的存在有可能是数据生成的错误或者恶意篡改的内容。也就是说，你要篡改成功，在比特币系统就必须改变系统中 51% 的副本（即 51% 攻击），由少数成为多数。可想而知，篡改成本从技术难度、时间消耗、人员使用上都是巨大的。

全网交易信息的绝对真实可靠、不可篡改、不会丢失，这种思想毫无疑问是当前所有信用监管模式的加强版和升级版。把区块链和现有的各种信用

评估渠道结合，充分发挥区块链这方面的优势，可以为建立真正有效的信用社会模式找到解决方案。

以客观真实的数据为基础，价值交换才有了可能。这是区块链在价值互联体系里的第四个特点。

5.2.5 可编程经济

可编程经济（Programmable Economy）作为一种基于自动化、数学算法的全新经济模式，把交易中的执行过程写入自动化的可编程语言，通过代码强制运行预先植入的指令，保证交易执行的自动性和完整性。这会带来前所未有的技术创新，在执行层面大大降低交易的监督成本，在减少造假、打击腐败和简化供应链交易等机会主义行为方面均有巨大的应用前景，是未来新经济的发展方向。

133

区块链的智能合约就是典型的可编程技术，不但可以自动实现预先设定好的智能运作，而且结合区块链真实、不可篡改、安全可靠的特点在经济生活中发挥巨大的作用。区块链可优化交易组织形式，有效促进经济社会运行效率的提升。

技术创新的变革力不同程度地影响了交易成本的变化。正是在这种成本结构急剧变化的条件下，依赖广域覆盖的互联网络的强力支撑，实体组织的边界在变小，虚拟组织的边界在变大，并呈现模糊化、开放化甚至消失的趋势，导致独自在区块链上运行的自治实体的出现，如分布式自治组织、分布式自治公司，这是人类追求经济效率所形成的新组织体。

从历史的角度来看，自治组织边界的变动不仅会改变其内部的连接结构，也会改变外部的连接状态。有目的的互动行为会促使组织自发地向更高级的

形式演进，从而改变了整个社会组织的连接结构和监管方式。区块链带来的交易成本节约和信任重构，既提高了社会管理效率，也完善了社会治理方式。区块链也许最终会带领人们走向更加公正、秩序和安全的自治社会。

区块链的智能合约和充分的可编程扩展性为适应未来智能社会的到来做了充分的准备，这是区块链在价值互联体系里的第五个特点。

5.2.6 区块链重构大数据产业

当今的大数据产业发展中还是存在明显的问题。

数据来源的参差不齐，不仅影响了数据挖掘的效果，也会造成数据分析结果的误导。而交易过程中由于数据的可复制性，交易中介方并不被信任，影响了分享和交易的意愿。但数据如不交易或者流通，让数据的产生、分析、利用仅停留在封闭环境之内，这样的数据孤岛将使数据的价值大大缩水。如何在保护数据隐私的前提下实现大数据的价值，更是当前的一大难题。

区块链技术可以充分改善大数据行业多个环节的诸多要素，不但赋予了大数据真实可靠的权属认定和方便的转移交换方式，而且确保了数据的高质量，为数据安全和隐私保护提供了完善的解决方案。

区块链技术为数据的共享交易提供了坚实的基础，大数据可以在区块链网络上便捷地交易，充分实现了数据的价值，并由此在经济规律的支配下实现远比现在更优化的社会配置和高效利用。

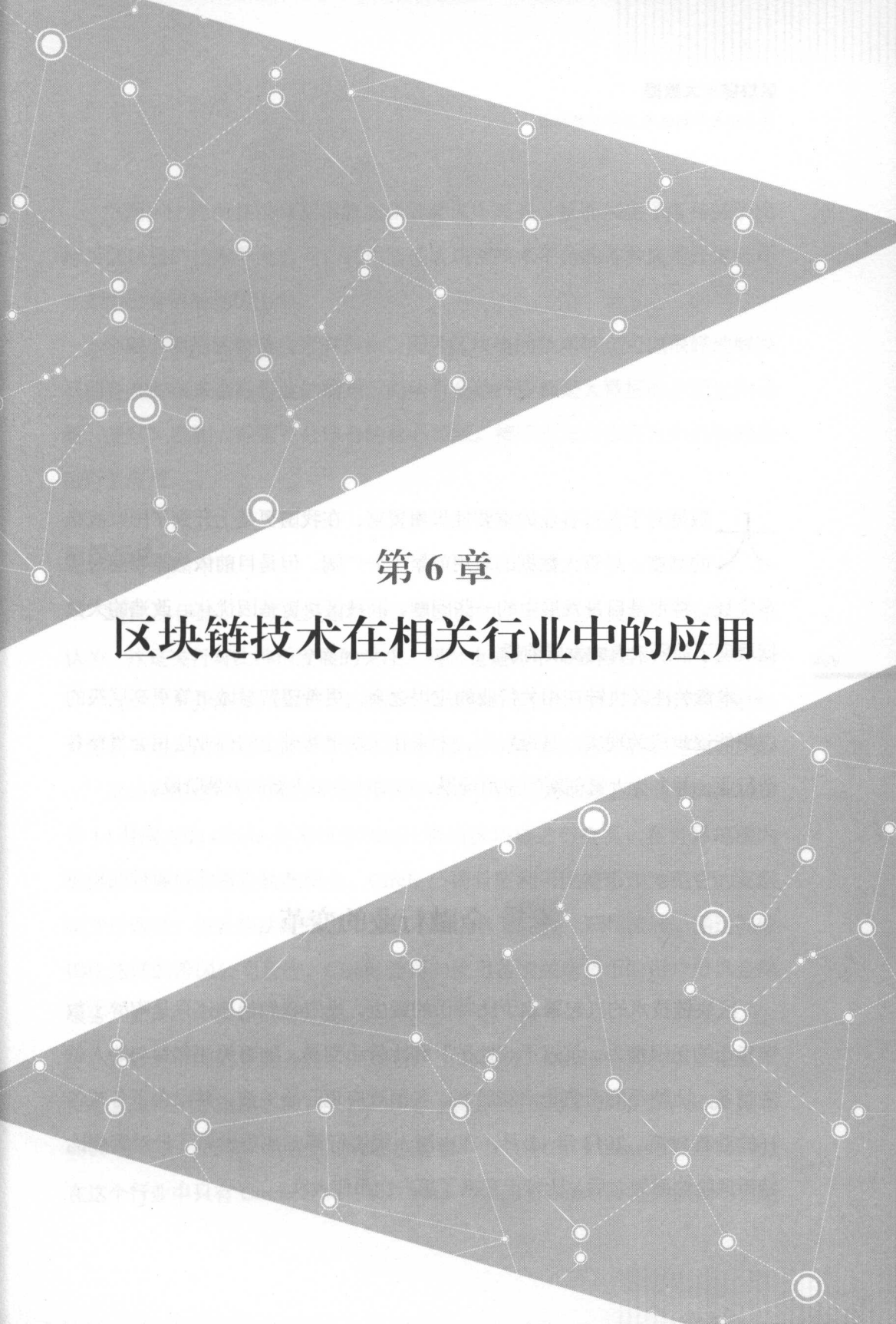
利用区块链的智能合约，可能实现更小粒度的数据交易模式，如条目交易、后付款的信用交易、充值交易、授权场景交易、数据交换交易等，从而改变当前大数据交易的商业模式。

依托区块链体系，能够建立针对所有用户的征信机制，推动全社会信用体系的建设上个台阶。

利用区块链技术脱敏后的数据交易流通，则有利于突破信息孤岛，建立数据横向流通机制，并基于区块链的价值转移网络逐步推动形成基于全球化的数据交易场景，建立真正属于全社会的大数据产业。

区块链技术在相关行业中的应用



The background of the page features a dark, textured triangular area in the top-left and bottom-right corners. These areas are filled with a network of white dashed lines connecting various-sized white circles, some of which are larger and more prominent than others, creating a complex, web-like structure. The central part of the page is white and contains the chapter title.

第 6 章

区块链技术在相关行业中的应用

大数据对于各行各业的重要性毋庸置疑，在我国更是上升到了国家战略的高度。尽管大数据的应用前景十分广阔，但是目前依然需要面对诸多挑战，特别是自身发展中的一些问题。而被区块链基因优化、改造的大数据展现了前所未有的发展前景。

138

本章关注区块链在相关行业的应用之余，更希望智慧城市等更高层级的应用能逐步成为现实。毫无疑问，未来在区块链基础上的商业应用会贯穿各个行业，并产生大量创新的应用场景，推动社会迈入新的发展阶段。

6.1 金融行业的变革

区块链技术的兴起肇启于比特币的诞生，投资者们起初还只是陶醉于数字货币的无限潜力，沉迷于“挖矿”和比特币交易。随着关注和参与的人越来越多，比特币的币值也水涨船高。各国政府也开始关注，并出台了各式各样的监管政策。2013年12月，在中国人民银行等五部委发布了针对警惕比特币风险的通知之后，比特币迎来了第一次币值大跌。

然而，比特币并没有在币值泡沫的破灭中死去，反而催生了各种基于比特币区块链的技术开发公司，利用其他区块链技术平台的各种应用开发公司也如雨后春笋般涌现出来。

金融方面的区块链应用有很多，因为区块链的技术特点可以很好地解决目前存在的很多金融行业的痛点。而所有金融行业都是大数据的主要应用场景，是区块链和大数据充分结合的核心领域。接下来重点讲述几个典型的金融行业应用。

跨国汇兑

2017年伊始，中国人民银行约谈了国内的三家比特币交易所。业内人士认为，这是央行对比特币发展的关注，同时也是为了加强监管。

也许有人会疑惑，比特币交易竟然会造成外汇的流失？其实道理很简单，只需要甲用人民币购买比特币，再以美元卖出，就可以完成货币兑换的过程。

正是利用这一点，很多创业公司开始在小额汇兑业务上施展拳脚。2013年10月成立的Circle公司便是利用比特币区块链进行交易，在世界范围内提供即时廉价的资金转账服务。Circle公司希望利用比特币作为免费的互联网支付网络，使各国法定货币在全球范围内毫无阻碍地顺畅流转。通过与像IDG这样的跨国公司合作，Circle公司向使用各国法定货币的用户提供金融服务，如人民币和日元的兑换。

凭借对传统汇兑业务在时间和费用上的优势，Circle公司的事业在资本市场上也得到了追捧。2014—2016年，他们分别获得了2700万、5000万和6000万美元的投资。而且，Circle也是第一个获得Bit License牌照的公司。在这个行业中只有Circle公司一家，Abra、Align Commerce这两家公司紧随

其后，大有颠覆行业之势。

跨国汇兑是最先成熟的区块链典型应用，实际上已经无故障地运行了数年，充分验证了区块链技术的完备性。正因为此项汇兑业务直击了社会的痛点，所以该应用的效益最明显，应用也最先成熟。基于区块链的跨国汇兑长期稳定的运行也进一步证明了区块链技术的巨大优势，为区块链的发展打开了新的天地。

银行

新兴的技术和挑战者的崛起使传统的银行业不能再墨守成规。2015年底，花旗银行出于降低自身运营成本的目的，发行了自己的数字密码货币——花旗币（CitiCoin）。很快，日本的三菱东京 UFJ 银行也于 2016 年初推出了他们的数字密码货币三菱币（MUFGCoin），并计划于 2017 年秋正式投入使用。

银行业决心利用区块链技术改造自身系统的态度已经逐渐明确，而银行间的交易也不是某家银行单独采取行动就能取得成功，这个时候就需要大家一齐协作，抱团取暖。R3 CEV 公司创立的 R3 联盟，迄今已经吸收了超过 50 家世界巨头银行和金融机构参加，其中也包括我国的多家银行。虽然联盟成员因为费用负担的原因发生了退出联盟的事情，但是退出者并非放弃区块链技术，而是改变了发展的策略。

正因为区块链的技术优势和实实在在的应用（汇兑），所有商业银行都在跟踪区块链技术的发展，并做了很多积极的尝试。虽然区块链参与商业银行业务的路线图尚不明朗，或者说有很多选项还在挑选中，但商业银行应用区块链技术将是确定性的事件。

中央银行层面也不甘人后。2017年2月4日有媒体报道,中国人民银行推动的基于区块链的数字票据交易平台测试成功。据了解,随后央行旗下的数字货币研究所也将正式挂牌。这意味着在全球范围内,中国人民银行将成为首个研究数字货币及真实应用的中央银行。

证券

纳斯达克交易所(NASDAQ)是证券交易市场最早吃螃蟹的人,它在2015年下半年就推出了新的针对一级市场的交易平台 Linq。该交易平台是基于比特币交易技术,用于一级市场公司的交易。Linq 拥有记录 IPO 前企业的股权交易,可以跟踪企业股权的交易、股东和投资人信息等,对私募市场的资产分析具有重要价值。

141

并非只有传统的股票交易所想着如何把股票交易放上区块链,某些电商平台也在做着同样的尝试。2016年12月,电商平台 Stockstock 在自己的区块链股票交易平台“TØ”(念 T fita)上发售自己公司的股票,成为了第一个在区块链上发售股票的公司。

除了传统的股票市场之外,众筹是一种新兴的募集资金的手段。区块链技术所具备的可信任特点使其在资产管理方面获得优势,因而能在众筹领域大有所为。Swarm 是一个基于 Counterparty 平台开发的众筹系统,它的底层是基于比特币基础协议的去中心化技术平台,具有提供交易和创建智能资产等功能。投资者可以在 Swarm 平台上认购初创公司的股份,并可以自由地转让。这个转让功能使投资者除了可以更灵活地退出之外,还可以在所投资的项目成功被巨头收购或者接受大笔投资时不会遭遇杯水车薪的问题。

6.2 智慧物联网

6.2.1 机遇

物联网（Internet of Things, IoT）作为互联网的进一步延伸，实现了物与物之间的直接联系。互联网技术连接的是全世界的计算机，实现了计算机背后的人与人之间的信息交流。物联网连接的是传感器、机器设备和各种控制中心，实现的是物与物之间的信息交互和控制。其不直接依赖人，却进一步扩充了人的能力，并将帮助人类迈向智能时代。

物联网是用信息构建起来的虚拟世界，丰富了物理世界的维度，推动了现实世界的发展。借助物联网，医院可以远程监控和应对病人的病情，工厂可以自动化生产，房间可以根据预先的设定调节室内温度和照明……这些信息再经过大数据的整合，可以产生全新的生产、生活、服务的维度，从而改造人类社会。

物联网虽然兴起比较晚，但发展很快。得益于高性能软硬件和信息技术的发展，低成本的物联网设备才能进入千家万户并呈几何级数增加。目前的物联网接入设备是 100 亿这个数量级，包括小型传感器及大型家电。预计用不了 10 年，物联网接入设备将达到 1000 亿的量级，并会相应产生海量数据。

6.2.2 现状与痛点

对于如此庞大的物联网网络和由此产生的海量数据，中心化数据中心的

基础设施投入、维护成本将会非常巨大，而且很可能还难以应付物联网数据的指数增长。大量的数据需要实时通过网络传递到数据中心并接受交互指令，无论是信息传送、处理，还是存储，都会面临持续增长的巨大压力。

在互联网数据安全频频出现问题的时候，更加深入人们生活隐私的物联网设备又进来了。你的居家生活习惯，家人的工作、学习和生活规律……这些数据如果都传输到管理中心数据库，又会产生怎样的安全问题呢？

如何有效应对海量的并且可能是非标准的数据，如何能够在物联网数据里保障数据安全和个人隐私、公司机密，这些都是物联网发展过程中必须面对和解决的问题，否则物联网的发展会很快碰到瓶颈。

目前的物联网架构基本都是封闭式的，虽然一个物联网系统的设备之间可以形成互联，并且也利用了互联网传输数据，但架构并不是开放式的，不同的物联网系统之间很难实现有价值的互联互通。这里面有一个很重要的原因，就是单个物联网的数据害怕被其他连接方非法篡改或者因交互导致数据丢失。

143

6.2.3 区块链能带来哪些改变

物联网面临的两个重大挑战——有效管理几何级数增长的设备和保障用户的数据安全，传统的解决方案会存在较大的困难，而区块链以其独特的技术特点可以战胜这些困难。

区块链技术为物联网提供了点对点直接互联的数据传输方式，让整个物联网解决方案不需要引入大型数据中心进行数据同步和管理控制，而由区块链网络自行完成分布式物联网的管理控制，并高效地进行数据保存。所有物联网信息都保存在区块链中，形成了可信的物联网数据来源。

区块链特有的数据加密保护和验证机制可靠地保护了数据的安全，并充

分维护了数据方的隐私或者机密。第三方即使侵入网络获得了存储的数据，也无法窃取真实数据内容，更无法对数据进行篡改和删除。所以，更多的小物联网可以通过区块链网络组成更广泛的、多维度的物联网，充分发挥大数据的巨大威力，而无需担心由于互联互通导致的数据泄密和损失。

另外，随比特币崛起的区块链，本身就有方便、可靠的支付结算潜质，为物联网平台带来了实时的交易结算能力。

6.2.4 具体应用方向

► 让物联网高效运行

当区块链应用于物联网时，区块链技术可以被应用于追踪过往的历史，也可以协调处理设备与设备、设备与人之间的数据交互和实际交易，赋予了物联网设备独立的身份。很多物联网设备，如智能家电，无需实时与云端服务器交互，专门构建云计算服务器任务来支持某个家电并无很大价值，特别是对于那些会在家庭存在十年甚至更长时间的设备。由区块链网络对数据进行处理和存储，并为系统内的物联网设备提供设备级别的控制和管理，会极大地降低物联网使用的成本，明显提高物联网系统的效率。

► 物联网的数据安全和隐私保护

物联网的数据安全和隐私保护问题越来越受到关注。由中心化服务器存储的物联网数据在存储、处理和传输环节有多种泄密可能，著名公司和政府部门的数据泄露事件层出不穷，让用户无法真正信任运营服务提供商的承诺。事实上，政府安全部门可以通过未经授权的方式对存储在中央服务器的数据内容进行审查，而运营商也很可能出于商业利益的考虑将用户的隐私数据出售给广告公司进行大数据分析，以实现针对用户行为和喜好的个性化推荐，

这些行为更是无法防止。

区块链技术为物联网提供了公开、透明、可追溯、不可篡改的数据保护措施，并通过特有的加密和分享机制保证了物联网数据使用的安全、便捷。

► 全新的商业模式

区块链可以确保物联网系统内部授权的真实可靠，从而为智能化的物联网设备赋予商业交易参与者的身份。这样，区块链全网获得交易身份的物联网设备就和人类一道，共同参与了区块链网络的交易。

例如，通过智能合约控制的冰箱可以在食品不够时直接向供应商下单进行采购，炉灶可以自动向燃气公司购买燃气指标……所有这些不仅需要支付能力，而且需要身份鉴权。

而在区块链网络上不同所有者的物联网设备所获得的加密数据，可以直接在区块链全网范围内进行结算交易，无需通过数据中心的再处理和交易撮合，不仅效率很高，而且也极大地节省了数据中心的存储和处理资源。

145

6.2.5 现有尝试

Filament

Filament 是一家区块链物联网初创公司，它在比特币区块链上建立了去中心化的物联网，为每个设备建立了身份，并创建一个智能设备目录，推进物联网设备之间进行安全沟通、执行智能合约以及发送小额交易等应用。

Filament 采用比特币区块链，使用包括区块命名和智能合约的协议，通过一个名为 Blockname 的协议创建一个设备的独特标识符，存储在每个设备的嵌入式芯片中，并记录在区块链上。当设备制造出来时就有全球唯一

的 IP 地址，其哈希值存储在区块链上。如果设备间需要通信，Blockname 会告诉它们对方的地址，Filament 可以让设备通过智能合约独立付款。

Filament 还开发了两个硬件模块，一个是 Filament Tap 传感器，允许与周边 10 英里以内的电话、平板电脑和计算机进行沟通；另一个叫 Filament Patch，可以实现硬件项目的定制。针对工业物联网市场，Filament 帮助石油、天然气、制造业和农业等行业建立设备之间去中心化的沟通平台，更好地管理生产作业。

IBM ADEPT 研究项目

IBM 是最早宣布区块链开发计划的公司之一，它在多个不同层面已经建立了多个合作伙伴关系进行尝试。

2015 年 1 月，IBM 宣布了去中心化的 P2P 自动遥测系统（Autonomous Decentralized Peer-to-Peer Telemetry, ADEPT）研究项目。ADEPT 平台由三个要素组成：以太坊、Telehash 和 BitTorrent。其中，Telehash 是一款使用 JSON 来共享分布式信息的私人信息传递协议，终端可以是设备、浏览器或移动应用；BitTorrent 是文件共享协议，用来移动数据，保证 ADEPT 的分散化特性。

IBM 还与三星专为下一代的物联网系统建立了一个概念证明型系统。该系统基于 IBM 的 ADEPT，使用该平台，两家公司都希望带来一个能自动检测问题、自动更新、不需要任何人为操作的设备，这些设备也将能够与其他附近的设备通信，以便于为电池供电和节约能量。

Slock.it

2015 年 6 月成立的 Slock.it 初步呈现了自治组织的运行状态。它是一个

建立在以太坊之上的物联网新共享经济平台，为全球首个分布式自治组织架构平台，欲取代 Airbnb、Uber 等作为中心调度站的角色，实现交易流程的完全自动化。

Slock.it 架构中的智能合约，可以保证用户在移动终端上随时随地追踪和控制出租或使用中的物品，每次共享完结时可以自动、准确、即时地收取费用，并且分配收入，给予分享者回报。

Ken Code 的 ePlug

ePlug 是 Ken Code 的一款产品。根据 Ken Code 的白皮书，ePlug 是一个小型电路板，位于里面的是“ePlug 认证”的电源插座和灯的开关。

为了安全性与可靠性，该产品提供了可选的 Meshnet (MeshNet 由 Mesh 与 InterNet 两个词合成，即在 Mesh 技术上建立的一种全新的去运营商的、真正无中心的分布式网络)、分布式计算、端到端的数据加密、无线连接、定时器、USB 接口、温度传感器、触觉传感器、光线和运动传感器，以及具有提醒功能的 LED 灯。

该产品以基于区块链的登录方式来确保安全。一旦输入正确的网络地址、URL，ePlug 所有者就会看到一个登录界面。

Tilepay

物付宝 Tilepay 是一个去中心化的支付系统，为现有的物联网行业提供一种人到机器或者机器到机器的支付解决方案。它基于比特币的区块链，且能被下载并安装到个人电脑、笔记本、平板或者手机上，所有物联网设计都会有一个独一无二的令牌，并用来通过区块链技术接收支付。

Tilepay 还将建立一个物联网数据交易市场，使用户可以购买物联网中各种设备和传感器上的数据，并以 P2P 的方式保证数据和支付的安全传输。

6.2.6 展望

物联网向区块链网络转型的发展速度可能会远远超过目前人们的普遍预期。有了区块链的帮助，物联网不但会即时开始其“智能化”的脚步，而且获得的物联网大数据也将会发挥更大的威力。当越来越多的物联网系统通过区块链网络连为一体时，物联网时代才会真正到来！

6.3 知识产权的管理

区块链技术正被应用于版权保护领域。通过发展基于区块链的数字版权管理（DRM）技术，对软件发行的每一份授权许可或者著作人对作品的版权进行记录和跟踪，使作者对自身的知识产权有更加强大的控制权。

6.3.1 机遇

2016 年，我国全年共受理专利申请 346.5 万件，同比增长 23.8%。其中，发明专利申请受理量为 133.9 万件，同比增长 21.5%；发明专利审结 67.5 万件……这还仅是我国一年的知识产权数据。如果从全球范围看，数量更是惊人：过去 50 年，仅全球的专利数量已经超过 1 亿条。这些海量的数据包含着丰富的知识产权信息，是一座巨大的“金矿”。

保护知识产权就是保护创新，用好知识产权就能激励创新。知识产权的

“十三五”规划已经明确被纳入国家“十三五”的重点专项规划之中，这也是知识产权规划第一次进入国家的重点专项规划，体现了政府对知识产权工作的高度重视和大力支持。但我国知识产权保护的法律环境与发达国家相比仍然有很大的差距，知识产权问题也一直是我国与发达国家贸易领域的谈判焦点之一。

知识产权数据资源汇集了人类智慧的结晶，对知识产权大数据进行有效利用和深度挖掘，不仅是推动科技创新和知识产权创造与保护的重要基础，而且能创造出巨大的经济社会财富。

知识产权大数据可以为创新主体提供多方面的重要服务。

(1) 找技术：创新主体可以用知识产权大数据发现自己想要进入的领域有哪些先进技术、最新研究成果是什么，以便于“站在巨人的肩膀上”创新。

(2) 找人才：创业创新需要寻找最契合的合伙团队和最优秀的研发团队，创新主体通过知识产权大数据分析，可以找到该领域最优秀的技术研发人员。

(3) 找商机：知识产权大数据可以帮助企业在专利运营过程中发现潜在买家，或者寻找合适的投资项目。

(4) 看趋势：知识产权大数据帮助企业了解产业最新发展态势，为企业制定商业决策提供支持。

(5) 避风险：分析知识产权大数据可以防范和规避知识产权侵权风险，为企业发展保驾护航。

知识产权大数据的巨大价值以及广阔的发展前景，吸引了越来越多的企业开始从事知识产权大数据的研发和应用。目前，许多从事知识产权服务的机构都很重视大数据资源的积累、建设和应用，开始尝试探索新的数据应用模式，并取得了一定的成果。但总体上，我国对知识产权大数据的挖掘还处

于起步阶段。

6.3.2 现状与痛点

按横向划分，知识产权产业可分为版权、商标、专利三个细分子行业。其中，版权的行业成熟度相对较高。而按纵向产业链划分，知识产权产业可分为确权、用权、维权三个环节。其中，确权是行业发展的源头，用权是行业内涵最丰富的环节，维权环节与法律制定与执行力度息息相关。

目前，知识产权产业痛点较多，按纵向产业链划分，主要可以归纳为三大方面。

（1）确权耗时长，时效性差。

知识产权的所有权注册，主要就是版权注册。理想条件下，这种注册不仅能精确地记录作品的原始所有权归属，还能记录所有涉及该作品的后续交易。

在中国，目前版权确权行业的效率在互联网企业的带动下提高了不少。较线下公司动辄几个月的时间不同，互联网确权服务公司基本能把时间控制在30个工作日内，一些网站还提供加急通道在10个工作日内完成确权过程。总体来说，互联网企业的进入加快了确权进程，但整体过程的时效性仍然较差。

在美国，尽管版权注册并不是必须进行的，但也确实提供了非常重要的法律优势，特别是在出现诉讼相关的纠纷中。虽然正式注册证书的颁发需要几个月时间，但注册的法律效力是从正式提交申请的日期开始生效，所以相对而言，时效性在美国不是大问题。

（2）用权变现难，供需失衡。

目前，知识产权行业存在严重的供需失衡，而且更多地体现在供需无法匹配上。从新增版权登记量来看，2015年我国的版权登记量（包括作品、软

件与质权)为164万件,年均复合增长达到18.7%,存量市场预计版权登记量最少在700万件左右。面对如此庞大且快速增长的版权供给,目前知识产权行业更应该考虑如何匹配供需,以促进自身的健康成长。

对于文学作家、摄影师和词曲作者而言,自身作品缺乏有效的证明和保护方法,传统的版权证明方式依赖权威的第三方认证,使用时成本很高,造成了内容流通和变现的困难。实际上侵权者有充分的途径绕开或者根本无视这些传统第三方认证渠道,而作者们也无法有效限制和打击侵权者,极大地制约了作者们对自有作品的使用。

在新兴的数字版权时代,如何有效解决版权保护的问题,方便版权交易和推广,确实需要有突破性的方法。

(3) 维权效率低,溯源困难。

从维权环节来看,追溯链条较长。

首先,界定侵权难度大,需要逐级查看授权说明才能最终确定侵权。特别是对于声音和图像这种数字内容而言,更容易引起所有权争议。这也是数字版权推广遇到的阻碍,因为很难分辨“原版”和“仿作”。

其次,维权追溯难,在确定侵权后,权利溯源难度较大。以音乐版权为例,词曲人拥有词曲的著作权,歌手拥有歌曲版权,此外还包括复制权、发行权、播放权、放映权等一系列的权利交叉在其中,权利归属的复杂程度可想而知。

最后,我国历史上版权违法现象较严重,下载盗版音乐、不标注引用来源很普遍,而这种现象似乎也反过来使用户对以前的侵权行为变得更加宽容。

6.3.3 区块链能带来哪些改变

当法律在为了知识产权保护而努力做各种尝试时,从提升效率的角度来

说，区块链技术也许能提供一个更好的解决办法。

(1) 确权：直接在区块链节点中声明所有权，理论上做到即时确权。

与传统模式相比，基于区块链的所有权确权有着独特的优势。

第一，注册过程几乎是即时的，并且收费也远比传统模式要低。

第二，区块链数据库的去中心化且加密安全性质，使作品不太容易遭受灾难性的损失或黑客攻击。

第三，作品的后续交易也会被实时记录，并且在网上可以被实时追踪到。

第四，区块链的公开透明可以更广泛地宣示作者对作品拥有的所有权。

虽然目前很多区块链注册缺少传统模式注册时所具有的法律效力，但是区块链注册已经可以作为传统注册方式的有效补充。一旦传统法律系统更好地接纳区块链注册，区块链记录就能够成为版权证明的强有力证据。

一个潜在的担忧在于，不同的区块链注册服务商使用不同的方法，可能会导致“多重注册”的冲突。这可能会影响区块链注册的权威性，并影响部分作者的注册意愿。不过，这个问题并非难以解决。鉴于区块链注册提供的不可改变的时间戳，平台之间或者公司之间可以相互承认，所有权纠纷就不难解决。

(2) 用权：点对点直接沟通，降低成本的同时加速供需匹配。

在明确权利主体的情况下，区块链技术还能帮助需求方与权利人建立点对点的直接沟通，减少中间繁杂环节。与大数据相结合，将会加速供需匹配和权利流转速度，大大降低中间成本，一举解决变现难题，提升变现效率。

通过智能合约还可以加速版权收费和交易的执行效率，创建分账的合同就是一种有效的方法。例如，当一首歌中，A 的贡献占 25%，B 的贡献占 75%，一旦 C 用户听了这首歌，那么系统就会自动为 A 和 B 分账。利用区块

链系统附带的数字密码货币，还可以便捷地实现小额实时支付。

(3) 维权：权利归属清晰，快速定位侵权行为与侵权主体。

知识产权维权难的关键在于，确定侵权困难，权利主体难寻。而区块链技术能通过程序算法自动记录信息和规则，具有明确、清晰的权利归属，能在第一时间确认侵权，同时也能快速找到侵权主体。由于侵权记录被不可更改地保存下来，权利人主张权益的成本大大降低，无需效率低下的第三方参与仲裁，从而解决了知识产权产业链冗长杂乱的问题，维权道路通畅，极大地提高了维权效率。

而且，侵权记录在全网都有体现，侵权者的信用会受影响（可借助大数据进行信用评分），不利于其以后的其他交易。因此，在基于区块链的知识产权体系下，失信者将越来越没有市场。传统市场中比较弱势的自然人、小微企业，将因为区块链所赋予的能力而实实在在地成为权利人。

153

6.3.4 具体应用方向

➤ 确权相关

版权注册：与传统版权注册形成双轨制

区块链版权注册可以从边缘作品切入，为大量未注册的作品提供方便、廉价的选择。作者们可以为作品创建一个时间戳，使其具有一种持久、稳固的区块链版权身份，伴随作品逐步走向盈利。对于边缘作品来讲，原本无人关注其版权属性，这部分商业价值获得了被发掘的机会。

这种双轨制甚至可能会产生新的模式。例如，传统版权注册可能最终会允许作品首先进行区块链注册，然后当作者需要获得额外的作品保护时再轻松地转入传统版权注册系统进行注册。

在先权利公证：提供“存在证明”

为区块标注时间戳是区块链独有的过程。通过区块链的时间戳生成“存在证明”，可以实现在不泄露文件内容的情况下证明其文件的所有权，特别是有时效性的所有权，如用于证明知识产权领域的在先权利。

匿名上传文件后，文件的哈希值会随之生成。这些哈希值将会成为交易的一部分，区块的时间戳就成为该文件的时间戳。但文件本身并不会存储在网络上，所以用户的文件内容没有被恶意公开的风险。区块链仅仅作为存储文件证明的公共账本，以备在以后面临所有权纠纷问题时的核查。

软件开发者可以使用这项服务来确认代码的版本，发明家可以使用这项服务来证明他们在特定时间就已经拥有某种想法，作家可以使用这项服务来保护他们的作品。即使黑客攻击了你的服务器，他也不能修改哈希值。也就是说，区块链技术可以从本质上证明这些数字资产的所有权。

➤ 用权相关

产权转移：去中介交易

艺术家们可以把作品（歌曲、画作、文章）放到区块链上声明版权，并进行点对点的自由交易，不再通过第三方组织和机构，无需任何中介。这种交易过程会大大提高效率并降低中间成本，特别是为边缘艺术家们提供了申请版权权益的机会。

版权印：让作品与版权不用分离

针对音乐等领域，一种更加激进的方式是基于版权印（Copyright Inside）的理念，让作品与版权不用分离。

版权印是利用区块链构建的一套结构化描述作品版权信息和授权规则的版权标识。通过版权印的投放、邀约、交换、披露和反向查询机制，可实

现多平台上交易达成和多平台间交易协作,使版权随时随处可查询、可交易。

权利人只要向版权印平台上传自己的作品,并自行选择授权范围、期限、价格,便自动生成一个含有版权授权信息的短链,也就是给作品盖上了版权印。用户只要点击作品的版权印,按选择的授权条件支付费用,就会自动生成授权书,迅速获得使用授权。

产权众筹

文字、音像等这类知识产权较难得到保护的产品一旦发布出去,很容易被盗版者窃取,一方面给版权作者造成经济上的损失,另一方面使产品品质在盗版传播过程中下降。

从经济学的角度来看,法律手段保护知识产权的成本较高,市场缺乏一种保护知识产权的正向激励。对于已经获得如文字、音像等知识产品的消费者,他们没有动力保护其已经购买的商品,而是宁愿把一本书或者一个专辑免费分享给身边的人。

借助基于区块链的合约币应用,用合约币加密或绑定的知识产权使用权唯一且不分离。这种使用权不能无偿分享而只能转让给他人,将知识产权及产品转让给他人的同时,自己便失去了对该知识产权的所有权和使用权。使用权不能共享有效防止了盗版,提高了品质。另外,通过溢价收入给先入者分红,可以给先入者足够的经济激励,从而保证他们传播正版的知识产权。

➤ 维权相关

所有权跟踪: 确认权利历史和归属

在知识产权保护领域,将贵重的作品所有权登记在基于区块链技术的公共账本上,被全网络验证并记录之后将无法修改。一旦产生产权纠纷,可追溯作品交易的历史,确认作品的所有权归属。

► 反版权

著佐权：互联网的版权观

著佐权（Copyleft）是一个由自由软件运动所发展的概念。在自由软件授权方式中增加著佐权条款之后，除了该自由软件允许使用者自由使用、散布、修改之外，还有著佐权条款要求使用者修改后的衍生作品必须以同等的授权方式释出以回馈社会。

著作权（Copyright）的版权理念强调产权，以确立私有产权的方式鼓励知识生产，限制知识的自由传播，“侵权必究”；而著佐权源于开源社区，强调共享，以产权共有的方式保护知识，限制知识的私有，“私有必究”。区块链的公开透明特性，使其同样十分适合著佐权。

6.3.5 现有尝试

将知识产权和区块链相结合是当前的热点之一，已有多家公司使用区块链技术来注册和阻止版权侵权。

Ascribe：艺术品数字产权

Ascribe 让艺术家们可以使用区块链技术来声明所有权，发行可编号、限量版的作品，可以针对任何类型艺术品的数字形式。它甚至还包括一个交易市场，艺术家们可以通过他们的网站进行买卖，而无需任何中介服务。Ascribe 还是最早通过支持 Creative Commons 许可（著佐权）的区块链服务。

Ascribe 可以在艺术家发布作品时确保对所有权的声明是安全的，买家可以收集并真正拥有这些数字作品。通过 Ascribe，买卖数字产权将会变得和交易实物产权一样简单。Ascribe 让你可以交易任何知识产权给任何对象，无论

他在地球的什么地方，只要通过电子邮件就可以做到。

BitProof：学历公证

BitProof 是近些年来涌现的众多文档时间戳应用中最先进的一种，将会让传统的公证方式成为过去。

BitProof 和一家旧金山的 IT 学校进行合作，把他们学生的学历证书都放在区块链上，重新定义了数字时代的文凭和学生证书的方式。

UProov：嵌入式权益证明

这款智能手机 App 同样是利用时间戳——在拍摄照片和视频的第一时间会创建一个独一无二的时间戳密钥，并且直接写入区块链中。这些写入的记录是完全不可更改的，哪怕一个像素的改变都会让记录无法对应而出错。这些写入区块链的时间戳将成为所记录照片或视频的铁证，而且可以在任何需要的时候用来鉴定。

所有数据由于都能被记录在私有或者公有的区块链上，因而可以作为保险索赔的证明或者作为犯罪证据。

Colu：数字资产交易

Colu 是首家允许其他企业发行数字资产的企业，他们可以将各种资产代币化。尽管免费的比特币钱包 COUNTERparty 也允许发行简单的代币，并且在其他钱包持有者之间进行交易，但 Colu 的代币可以设置成各种状态和类型，并且当在区块链上存储数据过大时能够将数据存储在 BitTorrent 的网络上。

Colu 允许使用比特币区块链在互联网上创建一种数字资产层。开发者完

全不需要了解比特币的相关信息，就可以为各种用途发行或者管理数字资产，从金融领域（如股票、债券、有价证券等）、记录（证书、版权、文档）到所有权（即使是门票、优惠券、礼品卡等）。

Colu 已经和德勤建立了合作伙伴关系，并且宣称已经被 20 多个系统所接受，Warranteer 就是其中之一。

Warranteer：云保修单

Warranteer 使用 Colu 让产品质量保修书从纸张形式变成区块链形式，随时确保它们是最新状态并容易转移。他们的客户已经包括 LG 和 GoPro。

Warranteer 领导的 eWarrantly 革命已经进入纯粹的数字时代，为保修书设立新的安全标准，并把所有的电子保修书保存在云端。

Wave：航运产业链

Wave 针对全球供应链，特别是在进口海运提单中出现的问题，使用区块链技术使其变得更加现代化。他们把供应链上的所有成员都放在去中心化的区块链上，允许成员之间直接交换信息和文件，解决了航运业的最大问题。Wave 的应用程序可以在区块链上管理文档的所有权，消除争议、伪造和不必要的风险。

Wave 已经创建了点对点的、完整的去中心化网络，包括货运商、银行、代理、贸易商以及全球贸易供应链中的各方都将会被连接到一起。通过使用去中心化技术，各方之间的通信都会是直接进行的，而不必通过一个特定的中心化机构。由于其天然的去中心化属性，Wave 网络将不会再有任何单点故障问题，也不再依赖任何单个实体。

Mediachain：数字媒体作品嵌入式版权

Mediachain 是总部位于纽约的区块链初创企业，要解决的问题是帮助数字媒体作品的作者将自己的作品打上版权标识。其初步设想是提供工具，让希望分享数字媒体的人找到自己喜欢的作者。

Mediachain 的核心产品是一个数据协议，内容创作者通过它可以给自己的作品附加信息，并把该数据打上时间戳放到比特币的区块链里，然后放到吸收了区块链技术的分布式文件系统（InterPlanetary File System, IPFS）上。这是一个整合了区块链技术各方面特征的分配式的文件系统。

Factom：公证防伪

159

公证防伪也是区块链可以大显身手的领域。区块链由于具有公开透明、安全性和唯一性的特点，因而天然具备了匹配公证防伪需求的条件。开发者能够创造新的应用程序，并把关键信息保存在区块链上，一方面可以确立关键信息的唯一性和真实性，另一方面可以维护可追溯的历史记录，实现公证防伪的各项需求。

Factom 维护了一个永久不可更改的、基于时间戳记录的区块链数据网络，大大减少了进行独立审计、管理真实记录、遵守政府监管条例的成本和难度。商业社会和政府部门可以利用 Factom 简化数据记录的管理，记录商业活动，并解决数据记录安全性和符合监管的问题。

法链：电子证据公证

大规模商用电子存证区块链联盟——法链宣告成立。法链是由 Onchain、

微软（中国）以及法大大等多个机构参与建立和运营的证据记录与保存系统，是一个多方参与的开放式区块链联盟。

通过法链上多方参与管理及运营，电子合同的签署时间、签署主体、文件哈希值等数字指纹信息会广播到法链所有成员的各自节点上，所有信息一经存储，任何一方都无法篡改，因而充分满足了电子证据司法存证的要求。

基于区块链技术中心化、防篡改、数据零丢失等特性，法链的建立使电子存证相关的法律服务变得更可信、高效、安全。

6.3.6 展望

效率至上，区块链未来将从本质上影响知识产权和法律领域。

区块链的不可更改特性能够让知识产权保护变得更加简单和低成本，通过完全自动的方式让每一个人都可以对任意的数据信息进行快捷注册和备案，这将会彻底改变目前全球知识产权保护的格局。因为就像当软件复制的边际成本接近于零时，也就意味着它和传统产品的发行会变得完全不一样。而知识产权注册的成本接近零时，有可能会诞生一个空前庞大的微知识产权交易市场，彼时的知识产权大数据将大有可为。

此外，通过技术手段，直接的“数字侵权”将很容易识别，但间接的“哈希问题”将成为巨大挑战。作品的哈希值与作品本身不是一回事，版权通常都比具体的一些二进制数排列要更广泛。对原作品进行轻微修改后的作品仍然受到原始作品版权的保护，但它们的哈希值是完全不同的。也许将来需要借助大数据和人工智能来判断两个作品的相似程度。

6.4 智慧医疗

6.4.1 机遇

近年来,很多国家都在积极推进医疗大数据的发展,不仅传统医疗机构在推行医疗信息化,制药企业也试图通过部署和使用大数据来节省高昂的研发成本。而互联网巨头们忙着对医疗领域收购投资,还有如雨后春笋般崛起的可穿戴设备制造商也在帮助人们利用大数据实现自我健康管理。医疗行业成为率先迈入大数据时代的传统行业之一。

早在2013年麦肯锡就认为,属于医疗行业的大数据革命即将到来,因为医疗行业早就遇到了海量数据和非结构化数据的挑战。

麦肯锡认为,医疗大数据逼近引爆点的合力来自以下四个方面。

(1) 需求侧:医疗行业对更好数据的需求。

在医疗改革和经济增长趋缓的大背景下,医疗成本的不断上升给各国政府带来了巨大压力。此外,过去的医疗系统更多是根据诊疗量而不是诊疗效果进行收费,医院和医生也就都没有大的驱动力去改善诊疗效果。所以,新的风险分担模型开始逐渐引入,以遏制费用和鼓励明智地使用资源。在新的机制下,医生的报酬和患者的治疗效果或总成本控制挂钩。

在临床领域,更多利益相关者开始接受循证医学的概念,医生给每位患者制定的医疗方案都要依据可获得的最佳科学证据。在许多情况下,将个人数据收集合并为大数据算法是最好的证据来源,这方面已经有大量的积极尝

试和成果。

(2) 供给侧：医疗数据的爆炸以及非医疗公众数据的聚合。

随着医疗信息化的不断深入，在临床领域，病人数据量呈指数级增长。2005年，只有30%的医生办公室和医院使用最基本的电子医疗记录。2011年底，这个数字上升到了50%以上的医生和近75%的医院。此外，约45%的美国医院现在或计划不久的将来会参与当地或区域医疗信息交流。

除了临床数据，大数据技术能让更多渠道的数据融入进来，包括索赔和成本数据、制药研发数据、患者行为和情绪数据等。与此同时，一些拥有大数据接入权限的公司和研究机构也正在通过合作或商业化的方式扩展其能力，进入其他数据领域。例如，全球十大制药公司联手形成的“TransCelerate Biopharma”合作，以简化和加速药物研发。

(3) 技术侧：分析技术和工具进步，推动信息共享。

技术的进步正在克服许多传统的技术障碍，电子病历系统比过去更便宜，数据交换也更简单、方便。除了促进纵向研究，技术的进步使清洗数据和保护病人的隐私也更容易。

通过获取个体日常行为数据，可能显示出一个患者最近户外活动较少，并在网上查找副作用的信息。这两点可能代表一种疾病的早期指标，需要进行早期干预以防止更严重的医疗事件。随着新数据互联互通和可以被使用，新的应用程序使数据更容易分享和分析。这些进步也开始提高医疗保健的质量和降低成本。

(4) 政府侧：政府对医疗大数据的推动。

各国政府都在推动医疗大数据领域的创新，不仅增加透明度，更可以帮助患者。例如，意大利药品机构收集并分析临床昂贵的新药作为国家成本效

益计划的一部分,并根据结果重新评估价格和市场准入条件;美国联邦政府一直在通过各种政策和措施鼓励医疗数据的使用,包括立法和激励以促进数据发布和可访问性、增强数据标准化和易用性、通过会议的形式宣传优秀的数据应用。

美国在1991年就推出了HIPPA(医疗电子交换法案),为病人的病例保密问题、网上就医系统的可靠性问题、急救调度响应系统的可用性问题等保驾护航。美国的医疗数据库信息也分为个人可识别信息和非个人可识别信息,前者设置严格的授权访问机制,后者则可以供医疗从业者拿来计算分析。

根据麦肯锡全球研究院的测算,仅在美国,如果医疗保健行业对大数据进行有效利用,就能把成本降低8%左右,从而每年创造出超过3000亿美元的产值。

中国对医疗行业大数据的发展也提出了要求。2016年,国务院办公厅印发了《关于促进和规范健康医疗大数据应用发展的指导意见》(以下简称《意见》),部署通过“互联网+健康医疗”探索服务新模式、培育发展新业态,努力建设人民满意的医疗卫生事业,为打造健康中国提供有力支撑。

《意见》指出,到2017年底,实现国家和省级人口健康信息平台以及全国各级药品招标采购业务应用平台互联互通,基本形成跨部门健康医疗数据资源共享共用格局;到2020年,建成国家医疗卫生信息分级开放应用平台,依托现有资源建成100个区域临床医学数据示范中心,基本实现城乡居民拥有规范化的电子健康档案和功能完备的健康卡,适应国情的健康医疗大数据应用发展模式基本建立,健康医疗大数据产业体系初步形成,人民群众得到更多实惠。

《意见》部署了14项重点任务和重大工程,主要包括:建设统一权威、

互联互通的人口健康信息平台；推动健康医疗大数据资源共享开放；推进健康医疗行业治理、临床和科研以及公共卫生的大数据应用；培育健康医疗大数据应用新业态；研究推广数字化健康医疗智能设备；发展智慧健康医疗便民惠民服务；全面建立远程医疗应用体系；推动健康医疗教育培训应用；推进网络可信体系建设；加强健康医疗数据安全保障；加强法规和标准体系以及健康医疗信息化复合型人才队伍建设等。

《意见》强调，要建立党委政府领导、多方参与、资源共享、协同推进的工作格局；要从人民群众迫切需求的领域入手，重点推进网上预约分诊、远程医疗和检查检验结果共享互认等便民惠民应用；支持发展医疗智能设备、智能可穿戴设备，加强疑难疾病等重点方面的研究；加快推进基本医保全国联网和异地就医结算；选择一批基础条件好、工作积极性高、隐私安全防范有保障的地区和领域开展健康医疗大数据应用试点；要研究从财税、投资、创新等方面制定政府支持政策，鼓励和引导社会资本参与健康医疗大数据的基础工程、应用开发和运营服务；要加快健康医疗数据安全体系建设，加强对涉及国家利益、公共安全、患者隐私、商业秘密等重要信息的保护。

6.4.2 现状与痛点

医疗健康大数据的想象空间很大，但距离实现还有一个漫长的过程。至少从目前来看，国内医疗机构之间信息数据互不流通的问题依然严重。大数据应用于医疗健康领域，面临的最大挑战是数据收集，其背后是数据隐私、数据安全以及数据可信度等一系列关键问题。

► 数据泄露问题

目前的医疗数据系统存在一个中心化的服务器当中。而现实中患者私密

信息泄露情况时有发生，医疗部门的中心化数据库或文件柜式管理已不再是最优选择。随着指纹数据应用和基因数据检测手段的普及，越来越多的人担心一旦发生泄露将会导致灾难性的后果。

随着医疗技术的发展，在病人身份背景、往期病史以及医疗支付情况的记录等方面，医疗数据正在起着越来越重要的作用。医疗数据可能是一个人的最隐私的数据，但由于网络操作错误或者黑客攻击等问题，这些个人隐私数据过去却存在大规模泄露的情况。

Anthem 泄露事件

美国第二大的医疗保险服务商 Anthem 公司信息系统被黑客攻破，近 8000 万员工和客户资料被盗。黑客入侵公司信息系统后，获得了公司员工和客户的个人资料，其中包括姓名、生日、医保 ID 号、社会保险号、住宅地址、电子邮箱、雇用情况以及收入数据。

UCLA Health 泄露事件

加州大学洛杉矶分校健康服务系统（UCLA Health）的 450 万份档案资料发生泄露。据悉，之所以此次泄露数据规模庞大，是因为它的客户数据（包括社会保险号，甚至包括诸如病况、药物治疗和化验结果的医疗数据）没有加密。

► 数据质量问题

医疗保健行业目前面临严重的数据质量问题，部分来源于医生的失误、黑客的攻击，部分来源于相同的电子健康病历（EHR）因为同时编辑而未能更新的问题。不同病例版本之间没有经过核对以及存在各种错误，使医疗

记录远没有达到可以被完全信任的地步。

► 数据分散问题

在医疗领域，每天全国数以万计的病人因看病检查产生了大量的信息数据，而且很多是比较私密的个人信息，需要被妥善保管。但在目前，病人信息基本由各家医院单独保存，或者卫生管理部门统一管理，导致储存的不确定性。

当真正需要使用病人的各种数据时，收集分散在不同数据库中的信息将带来大量时间和金钱成本。更多时候，各数据方的顾虑导致数据无法进行整合，医疗数据资产的外部性无法被充分发挥，整个产业的效率被大大降低。

6.4.3 区块链能带来哪些改变

现在来看，除了金融领域，医疗保健似乎是从区块链技术中获取最大收益的行业。区块链在医疗健康行业落地应用的关键不是性能，而是保障信息数据的安全和隐私。

(1) 保障数据安全和隐私。

医疗保健部门保存了大量的隐私信息，如病史记录、疾病、支付和治疗。在任何时候，对这些信息都应予以保密，但集中的数据库不再是一个切实可行的好选择。在上述数据泄露的案例中，往往是由于网络操作的问题所引起，使所有数据暴露在黑客的面前，一个单点故障就能够导致所有人的信息受到威胁。

数据的隐私问题在区块链架构上能够得到更好的解决。区块链由于高冗余、无法篡改、低成本和能进行多签名复杂权限的管理，也许是目前人类能找到的最佳数据保管方案。通过多签名私钥和加密技术，当数据通过哈希算法处理并被放置在区块链上后，能够只允许那些获得授权的人才可以对数据进行访问。使用这种技术，将能够依照一定的规则来对数据进行访问，无论

是医生、护士还是病人本身都需要获得许可。

(2) 保障数据不被篡改。

由于缺乏信息交互的分散系统中存在文件错位和被篡改的情况,运用区块链技术可以安全、准确地永久储存文件和保护客户记录。将病人信息以加密方式连接到区块链中,既能保证数据不被篡改,又可通过设置多把私钥来更加安全地保护病人隐私。

区块链不再由某单一团体控制医疗数据,而是让所有参与者共同负责数据的安全性和真实性。这就为人们提供了唯一共有的医疗保健数据来源,不再受人为错误和手动数据兼容性的影响。

(3) 优化流程和提高效率。

由于医疗收费系统过于复杂,政府和医疗机构每年要花费大量人力、物力维护该系统。如果保险公司、医院收费部门、贷款方以及患者都使用同一个区块链来管理支付,既能够保护患者的隐私,又能够提高医疗收费过程的效率。区块链的稳定性可以让所有相关方迅速地访问、查看和获取不依赖第三方存在的分布式总账,促进医疗数据安全地在机构间流动。

区块链在医疗健康领域的另一项重要应用是杜绝骗保,防止内外勾结,干预合约条款以及执行。保险公司可以获得真实可靠的医疗数据,患者可以快速地完成出险流程,医疗机构等相关方可以简化提供的各种手续和辅助工作,智能合约的执行会大大优化各方面的用户体验。

6.4.4 具体应用方向

从某种意义上讲,区块链是最有可能运用到帮助解决B端用户医疗信任的方式。

► 电子健康病历 (EHR): 以患者为中心

在医疗方面,区块链最主要的应用是对个人医疗记录的保存,可以理解为区块链上以患者为中心的电子病历。如果把病历想象成一个账本,原本它是掌握在各个医院手上的,患者自己并不掌握,所以病人就没有办法获得自己的医疗记录和历史情况,令其他医生无法详尽了解到患者的病史记录。但现在如果可以用区块链技术来进行保存,就有了个人医疗的历史数据,无论是就医还是自我健康规划都有历史数据可供使用。并且,这个数据真正的掌握者是患者自己,而不是某个医院或第三方机构。

► DNA 钱包: 安全分享数据

基因和医疗数据能够运用区块链技术进行安全存储,并且通过使用私人密钥来获得,这将形成一个 DNA 钱包。这使医疗健康服务商能够安全地分享和统计病人数据,帮助制药企业更有效率地研发药物——这种模式正在逐步建立起来。

► 数字密码货币支付: 更多隐私权利

以区块链技术为基础的、以比特币为代表的数字密码货币支付,可以给予病人更多的选择。这么做的主要原因是想利用比特币等数字密码货币的匿名性保护客户隐私,这一点是信用卡无法做到的。这样可以让客户在询问病情相关问题时有较少的顾虑,避免尴尬,特别是面对一些敏感的疾病时。

► 药品防伪: 审计追踪记录

与编码防伪技术类似的是,对于运用区块链技术防伪的药品而言,在药品包装盒表面有一个可以被刮去的面,底下是一个特别的验证标签,这可以与区块链相互对照来确保药品的合法性。

► 产业整合: 提升产业效率

产业链上下游所有参与者共同负责数据的安全性和真实性。例如,医院

收费部门、制药企业、贷款方、患者以及保险公司都使用同一个区块链来管理支付，既能够保护患者的隐私，又能够提高医疗收费过程的效率，整个行业中的冗余将大规模减少。

6.4.5 现有尝试

医疗行业也对区块链技术表现出了兴趣，很多企业如 DeepMind、Gem Health、Factom、BitHealth 等都在这个领域进行了尝试。

DeepMind：用区块链技术来更好地保护患者数据

DeepMind 是由人工智能兼神经科学家 Demis Hassabis 等人联合创立的前沿人工智能企业，其将机器学习和系统神经科学的最先进技术结合起来，建立强大的通用学习算法。2014 年 1 月，DeepMind 由谷歌斥资 4 亿美元收购。因为 AlphaGo 与李世石的“人机大战”，DeepMind 公司被广为熟知，并已经成为 AI 领域的明星。

DeepMind 欲将其算法应用到医疗保健行业，包括计划在 5 年内使用机器学习处理英国国家医疗服务体系（NHS）的数据。双方签署了一项为期 5 年的协议。然而，这也引发了针对患者隐私数据的争论。

医疗数据让 DeepMind 面临监管风险和反托拉斯问题，他们于是开始认真考虑采用区块链技术。如果用户通过区块链能真正拥有其医疗数据并控制授权范围，那么 DeepMind 便可以清晰地告知消费者和监管者：“用户实际拥有他们自己的数据，我们只是使用者。”这可能是 DeepMind 在其医疗区块链项目中的策略。

同样的方式也适用于基因测序领域。区块链测序可以利用私钥限制访问权限，从而规避法律对个人获取基因数据的限制问题，并且利用分布式计算

资源低成本完成测序服务。区块链的安全性让测序成为工业化的解决方案，从而推进数据的海量增长。

Gem Health：打造医疗健康产业的新框架

Gem Health 为医疗保健服务商提供网络基础设施，致力于构建一个全球化医疗保健综合体，并为人们提供更加私人化以及更低费用的服务。在医疗数据、理赔过程、药品供应链、健康网络、通用健康身份以及基因数据管理等方面，Gem Health 正在探索区块链技术的应用。

Gem Health 的创始人兼 CEO 米卡尔·温克斯佩克特（Micah Winkelspecht）表示：“区块链技术将会创建一个连接医药健康产业的新框架，将所有医疗平台的重要数据连接到一起。区块链技术保证了数据的有效性和安全性。对于医疗行业而言，这使医院、保险公司和实验室能够实时连接，并且即时无缝分享信息，而无需担心信息被泄露或者被篡改。区块链能够建立一个数据记录和身份管理的公开标准。一个全球化的医疗健康区块链能够将每个病人包含本地医院和医生的记录信息关联匹配一个 ID。基于区块链技术的通用医疗健康 ID 能够减少患者诊疗过程中的医疗错误，并保护病人隐私。”

区块链能够让其他多个组织来访问网络，而不必担心数据的安全性和完整性。病历可以被多方进行创建、共享，并且能够让多方进行追加更新，这将会重塑整个行业的效率和透明度。一个统一的区块链生态系统将实现公司与用户之间更简单、更安全的交互，实现跨界应用程序的配备。

Factom：保护医疗记录以及追踪账目

Factom 是一家区块链技术服务公司，维护了一个永不可更改的、基于时

时间戳记录的区块链数据网络，并利用区块链技术助力各种应用程序的开发，包括审计系统、医疗信息记录、供应链管理、投票系统、财产契据、法律应用以及金融系统等。

2015年4月，Factom 宣称与医疗记录和服务方案供应商 Healthnautica 展开合作，研究运用区块链技术保护医疗记录以及追踪账目，为医疗记录公司提供防篡改数据管理，通过运用 Factom 不可篡改的账本来对医疗记录和合约进行验证并打上时间戳，Healthnautica 的客户（医患双方）都希望能够提升赔偿流程的处理效率，确保医疗记录的真实性和有效性。

Factom 的技术首先将私人医疗数据进行加密编码，然后生成一个数据指纹以供打时间戳和验证。这种方式可以确保实际医疗数据不会泄露给第三方，保护了病人的隐私；接下来可以推动保险公司运用区块链技术保存数据，让各方能够即时验证合约有效性并提升医疗账单支付效率。

171

Block Verify：将区块链防伪验证应用于医药企业

Block Verify 是一家基于区块链技术的防伪方案服务商，目前能够鉴别的类目包括伪造品、调换品、被偷商品、虚假交易。除了奢侈品、宝石、电子产品之外，该公司的服务还可以用于药品追踪。

众所周知，药品的质量是否能够保障，真伪是否能够鉴别，异常重要。但从药品生产制作到分销商，再到各大医疗机构及患者手上，其中流通环节过多，如何把控成为了首要问题。

Block Verify 用区块链技术解决了这个症结。与编码防伪技术类似，运用区块链技术防伪的药品，在其包装盒表面内嵌一个特别的验证标签，应用区块链记录不可篡改、共享安全可靠的属性，可以有效确保药物的合法性和

真实性。区块链在药品防伪方面，不仅能实现真伪验证和正品保证这类较浅显的应用，还可以实现对药物的溯源追踪，甚至提供全球药物商品解决方案。药品被制造出后，通过登记，公司能够对药物流通的全部供应链环节实现实时监督，转换成为完整的药物质量监察解决方案。

6.4.6 展望

区块链和大数据是未来数字社会的基础设施，具有广阔的应用前景，但其对于现有医疗秩序规则及信息化系统的替代也不是一朝一夕之事。

在整个医疗健康产业里，传统医疗机构手握核心数据资源，可穿戴设备采集数据可以作为补充，但目前仍缺乏统一的数据接入平台。对于中国而言，破除内部、外部、公共和商业之间的壁垒成为首要解决的难题。将不同渠道的信息内容整合在一起，用区块链作为统一的标准去建立共享数据库，为大数据的计算分析提供信息池，需要有区块链专家、医疗专家以及 IT 专家等各方通力协作，才能有效推动区块链技术在医疗行业的发展及运用。

6.5 智慧能源

6.5.1 机遇

从钻木取火到蒸汽机，从发电技术到信息化，人类一直在探求更好地获取与使用能源的方法。能源流和信息流的融合正在推动新一轮能源革命。煤炭、石油、天然气都通过市场机制实现了互联，反而是最标准、最容易联网

的电能迟迟未有实质性的动作。

人类对能源的使用需求可以根据应用区域分为两类：第一类是大规模、大范围的消费需求，例如，城市能源必须通过大能源系统来提供，这就是长距离的能源输送；第二类是张扬个性、特立独行的需求，可以通过“分布式能源+区块链”来满足这些人的需要。

从技术层面来看，能源网络系统也在朝两个方向发展：一个是越来越复杂的、庞大的能源系统，例如，大电网从地市级到省级，再到区域电网、全国联网、洲际联网，最后发展到全球联网；另一个就是分布式能源系统解决就近消费的个性需求，一个个很小的能源区块链构建成了庞大能源系统的另一个属性。

据高盛公司预计，美国国家电网在未来数十年会向融合更多去中心化资源、实时报价系统和更紧密匹配需求以及供应的方向进化。进化的核心是通过智能电表、智能装备、可再生能源和能源储存的结合实现电网的现代化，这将催生一个价值 20 亿 ~ 70 亿美元的美国分布式能源市场。

173

6.5.2 现状与痛点

► 能源损耗大

传统的电力市场是由公共事业公司来垄断。电力产自于大型电站（有很多就建在能源产区，减少能源运输成本），远离人口中心，经过无数的输配电基础设施到达终端消费者。尽管中国在特高压远距离送电方面取得了长足的进步，但建设大量基础设施的成本和送电的损耗仍然是巨大的。中心化电站和终端消费者之间的电力传输损耗占比为 9%。

► 可靠性

中心化电网等基础设施存在由于可靠性问题而导致大量人口断电的风险，

维护中心化电网的安全运行也需要相当的成本。

► 负载平衡

传统中心化电网的用电负载存在明显的峰谷效应，对发电、输电、配电、储电等环节都提出了很高的要求。中心化的供电体系需要为负载平衡付出很多额外的努力。

► 无法交易

我国已有很多分布式的可再生能源，但面临一些消纳和高效发展问题。例如，安装了太阳能的用户多余的发电无法转让给其他用户使用，造成能源浪费。

6.5.3 区块链能带来哪些改变

► 去中心化的能源交易

区块链赋予去中心化的能源交易能力，结合区块链和通信技术可以促进数百万的参与者之间更安全地交易和支付。通过连接本地的能源生产者（如有太阳能板的邻居）与本地的消费者，区块链使分布式的实时能源交易市场成为可能。

区块链发挥准确、灵活、开放、去中心化的特点，通过设定交易机制（智能合约）把各个方面的利益相关方联系起来，如储能电池制造商、券商银行、分布式发电售电商等，让大家找到一种共同认可的合作方式进行协同创新。这就相当于手机只是提供高性能的平台和底层系统，为不同的开发商提供丰富的API和接口，让产业链上的各方都能够依托这个平台，快速开发出各种应用。

► 减少损耗，更加安全

区块链将通过一个去中心化的实时能源市场连接本地生产者和消费者，

减少长程传输需求，同时也减少中心化供应模型的固有缺陷。

电网现代化（由智能电表/装备、可再生能源和储存技术驱动）已经开始撼动传统的公用事业模型。区块链可以促进电网和新能源的发展，最终创造一个更加去中心化的电网。这样，拥有可再生能源的电力用户也可以同时成为电力生产者和电力市场的交易者。

一个区块链驱动的市场也能增强电网安全性，刺激智能电网科技的应用。以自产新能源为基础、以电网调节为补充的现代化用电模式，不仅更加绿色环保，而且有更高的安全可靠性能。

消费者自产能源的方式为整个电网的负载平衡提供了更多的解决方案。电网方面可以通过经济激励的方式调动不同消费者自产能源的总量，从而更好地实现负载平衡。

175

► 推动行业变革

（1）结构性影响：区块链技术会驱动更多分布式电网基础设施。

能够作为本地生产者在能源市场上进行交易这一点，会吸引更多资源投入赋能分布式电网的技术，包括智能电网装备、物联网装备和电动汽车。直接参与能源销售会带来更多的装机量，也就带来更多可输出电力。

电网越是分布式的，就越是能更可靠、高效地匹配能源供需，包括但不限于发送实时报价信息和减少昂贵的输配电基础设施开支。

（2）政策影响：区块链技术会终结净计量电价机制。

正是净计量电价这样的政策促进了分布式太阳能的市场接受度，美国越来越多的州里，这样的政策促成人们转向太阳能而非继续为电网付费。分布式的能源生产者会接受向电网售回电力的替代方案，也就是售回给其他本地的消费者，区块链网络可以提供这样的去中心化的安全交易市场。

6.5.4 具体应用方向

未来的电网将由亿万交互的终端组成,包括微电网、光伏系统、智能设备、分布式计算系统与能源管理软件等。

区块链在构建下一代分布式微电网体系中的潜力巨大,会产生数千万甚至上亿个去中心化节点,这些节点不仅能够交换信息,还能完成能源交易。区块链会带来一个去中心化的能源市场,不仅极大地推动分布式能源方面的投资活动,还会对电力市场收入进行再分配,传统中心化发电、输电和配电方面的部分利益会分配给消费者。

结合区块链技术与物联网技术,将使分布式能源交易的协商机制成为可能。分布式能源生产者可以自动广播盈余电量及其持续时间的相关信息,而消费者可以根据他们的能源需求进行自动响应。使用区块链账本,生产者和消费者的代理机器可以协商价格,达成能源销售交易。

传统电网的核心就是中心化的电力生产机制和长程输电基础设施。分布式资源,特别是屋顶太阳能,可以按净计量电价将多余电力有效地售回给电网,售回电网的报酬形式是净计量电价,用户可以因此抵扣掉一部分传统电费。

6.5.5 现有尝试

位于纽约布鲁克林的初创企业 LO3 能源 (LO3 Energy) 已经启动了区块链在电网的应用试验,项目的名称叫作“TransActive Grid”。这个项目构建在以太坊上的智能微电网交易系统,实现点对点的能源交易和控制。他们已经建立了双节点的模型,在微电网中收集消费和发电数据,并将其存储到区块链中。TransActive Grid 还包括智能仪表硬件层和使用区块链智能合约的软

件层,整个平台基于以太坊区块链构建,能够提供一个可审计的、无法篡改的、加密的、自动交易的历史,并自动执行智能合约。参与的家庭都配备了连接到区块链的智能仪器,追踪记录家庭使用的电量,并管理与邻居之间的电力交易。

TransActive Grid 的应用场景是在总统街道的一边 5 户家庭通过太阳能板发电,在街道另一边的 5 户家庭可以购买对面家庭多余的电力。而连接这项交易的就是区块链网络,几乎不需要人员参与就可以管理记录交易。

TransActive Grid 也是整个布鲁克林微电网的一个组成部分。2016 年 4 月,首笔基于以太坊的电力交易完成,布鲁克林的居民艾瑞克·弗儒明(Eric Frumin)把自己的太阳能电池板产生的多余电能直接卖给了他的邻居鲍勃·索凯利(Bob Sauchelli)。艾瑞克·弗儒明拥有的每单位能源被计算并记录在以太坊区块链中,然后使用可编程的智能合约指令使这些能源能在公开市场出售。如果邻居没有购买这些电能,产生的多余能源就以批发价格卖回电力公司。

通过这个实验,LO3 能源希望证明一个本地的智能能源交易系统要比传统的自上而下的能源配电系统更有效,也更节约成本;区块链对于构建微电网系统的潜力和价值也是巨大的。

Grid Singularity 是一家成立于维也纳的初创公司,它也在探索用区块链技术来实现能源的交易验证。该公司专注于在发展中国家的这一业务,希望使“即付即用”的光伏系统更加可靠。其终极目标是为能源系统创建一个区块链平台,拥有电网上所有类型的交易。

当然,区块链在电网的应用很大程度上还处于理论研究和概念验证阶段,但应用模式的探索更关键,这些早期的能源和用电设备验证将为未来区块链

电网积累更多的经验。

中国国家应对气候变化战略研究和国际合作中心主任李俊峰表示，对于能源互联而言，区块链就是一种机制、一个解决方案，未来会不断演进。当下最重要的是要推出促进分布式能源发展的产品，才能逐渐走向成熟。

6.6 供应链

6.6.1 机遇

在全球化大分工和信息化、智能化日渐深入的今天，社会供应链也变得越来越复杂，已经突破了传统的行业边界、企业边界，改变了传统的商业运作模式。崛起的大数据不但可以从供应链获得高质量的数据源，而且可以为供应链的管理和高效运行提供升级换代的核心驱动力。

由于数据的电子化和互联网化，生产现场的仪器、供应链渠道的各种物联网传感器和各种电商销售数据都可以作为大数据的高质量数据源。通过对供应链各流程环节的数据相关性分析，可以精确把握各环节的具体需求，成为供应链顺利运行的可靠保障。灵敏的销售预测可以优化库存策略，生产制造环节的安排和物流管理直接影响对客户的订单交付，而客户的需求可以直接反馈给生产设计的优化。大数据通过有效的、定性和定量的预测分析手段和模型，并结合历史需求数据和安全库存水平，可以综合、精确地指导供应链各环节，实现效率的提升和风险预警，带来更佳的供应链管理。

供应链大数据的本质是希望消除供应链各环节之间的不信任,包括供应商、制造商、物流、销售等方面之间的摩擦成本。多方数据库信息交互,共同维护供应链快速、准确的判断和反应能力,提升运作效率。大数据的数学模型、优化和模拟技术为复杂的生产与供应问题找到了优化解决方案,惠及参与各方。

除了流畅的供应链供给,提高供应链资金运作的效力,降低供应链整体的管理成本,也是当前各方对供应链领域发展的一个迫切需求。供应链金融(Supply Chain Finance, SCF)是商业银行信贷业务的一个专业领域,也是企业尤其是中小企业的一种融资渠道。

供应链金融的最大创新在于填补8%~20%的融资利率空白。假设一家供应商(中小企业)拿到了一笔大额订单,三个月之后才能获得客户回款,这三个月的资金周转困难如何解决?银行不愿意承受过高风险,不愿意给中小企业贷款。所以,此时供应商一般会去找保理公司,但扣除利息之后能换到90%的现金就算不错了。还有一些供应商被迫向民间借贷,利率更要高出不少,这对供应商而言无疑增加了生产成本。

中小企业由于信用和风险识别成本高,很难从银行取得资金支持。于是,赊账销售已经成为此类企业最关注的支付方式。这样供应链金融的价值也就逐渐被挖掘出来。

6.6.2 现状与痛点

大数据的概念超出了传统供应链数据产生、获取、转换、应用分析和存储的概念,不但有非结构化数据,而且数据内容也趋向多样化。这里不仅需要从不同的数据源头获得广泛的数据,而且需要对多样化的、非结构化的数

据建立强大的筛选和处理能力，并有能力面对海量增长的数据量的挑战。显然，这已经超出了传统 ERP 结构的处理能力，需要真正引入大数据的理念和方法，才能真正应对挑战。

在供应链链条中，核心企业及上下游企业的信息流、物流、资金流信息的整合至关重要。而这中间面临的挑战还包括第三方是否愿意全面地提供全部的数据，无保留地加入供应链大数据。目前的现状是各企业自己维护自己的数据信息，不愿意和上下游分享，这种信息孤岛增大了信息整合的难度。能否有效突破供应链各环节的数据孤岛，建立统一的供应链大数据，非常关键。

供应链中商品从卖家到买家的同时会伴随着货币支付活动。在高信贷成本和企业现金流需求的背景下，金融服务公司提供商品转移和货款支付保障，通常被称为供应链融资或者贸易融资。银行常见的贸易融资业务是保理业务，通常是银行在买家支付货款前代付货款，当然也会收取费用，但是仍然存在买家延迟甚至拒绝付款的风险。商品供应商需要负担商品保护和保理业务的高额成本，大大降低了利润，给货款支付带来不安全因素。另外，保理业务还潜藏了很多风险因素，包括拒付、重复支付甚至诈骗。除此之外，人工操作、纸质文件、汇兑等也增加了操作成本。

6.6.3 区块链能带来哪些改变

区块链的出现如同给供应链增添了润滑剂，提供了解决数据孤岛的解决方案，改善了供应链数据的共享和流动。

区块链技术具有数据不可篡改的特点，用时间戳的方式证明数据的存在状态，确保数据在交易各方之间公开透明并可追溯。同时，区块链特殊的隐私安全保护机制彻底根除了供应链各个环节信息共享中的障碍。区块链重新

塑造了供应链的信用体系，增加了交易各方之间的信任，同时降低交易成本，从而解决了供应链融资中的信用风险问题。

而且，区块链确保了资产贸易的真实性。区块链网络全方位、全透明地显示完整的供应链，将网络的每一个节点、每一份资产以数字化的形式在网络上展示。因此，区块链上任何节点的交易都会被全网认定，物流信息也可通过产品地理位置信息的改变在网络中体现，从而保证了数据的安全性和不可被篡改。供应链上的所有企业，特别是金融服务企业，可以依据区块链网络展示的真实、不可篡改的供应链信息做相应的金融服务。这样不但实质性地降低了自身的风险，而且降低了供应链企业获得服务的门槛和难度，降低了企业的成本，真正促进了供应链的健康发展。

6.6.4 应用方向和尝试

181

CargoChain

CargoChain 结合了物流 RFID 技术和区块链技术，能够提供货物的实时状态查询，并增加了运营相关方之间的信任。银行和保险公司可以监控不可伪造的数据，从而提供及时的金融服务。由于有清晰的、可验证的托管链，海关可以更快地清关放行，而进口商、出口商和生产商也就可以方便地获得相应的服务。

Fluent

Fluent 成立于 2014 年，为银行、企业和其他非银行客户提供多种基于区块链的服务。

供应商可以用 **Fluent** 网络的专用区块链技术来改善现有贸易融资。**Fluent** 开发的区块链网络实现了发票的代币化，不但消除了重复和伪造发票的风险，而且生成了可交易的数字资产。投资者可以简单地在 **Fluent** 平台完成支付，参与数字资产的交易。票据和应付款项可以以不同的方式分割和出售，盘活资产，获得方便的贴现。无需跟踪和协调汇款信息，区块链可以完全原样捕捉发票的信息，便利买家支付，不需要知道谁持有发票。

Skuchain

Skuchain 是一家位于美国加利福尼亚州的创业公司，致力于开发 B2B 贸易和供应链融资区块链应用程序。他们正在开发区块链解决方案，以此来解决价值 18 万亿美元的全球贸易金融市场仍旧依赖纸质文件的问题。

Skuchain 致力于解决商品流动和资金流动没有同步的问题，用区块链技术进行整合。最直接的应用场景就是消除假货以及跟踪物流方向，还可以通过打造无纸化的交易环节来释放信用证的交易成本。

6.6.5 展望

据麦肯锡统计测算，区块链技术在全球范围供应链金融业务中的应用，能帮助银行每年缩减运营成本约 135 亿 ~ 150 亿美元，减少风险成本 11 亿 ~ 16 亿美元。应用区块链供应链网络的买卖双方企业，一年预计能降低资金成本约 11 亿 ~ 13 亿美元，减少运营成本 16 亿 ~ 21 亿美元。

区块链技术有可能会成功改变供应链，并颠覆传统的商品生产、市场推广、购买和消费的方式。从长远意义来看，提高供应链的透明度、可追溯性

和安全性，能够促进一种信任和诚信的环境氛围，防止供应链系统中的一些不规范做法，从而有助于构建更安全、更可靠的经济体系。

但是，供应链金融 + 区块链的发展不会一蹴而就，需要适应整个大环境的变革。区块链技术还在不断成熟的发展过程中，无论底层技术、应用开发、市场培育、商业推广，还是社会认可，都需要比较漫长的过程。而大数据在面临自身固有的瓶颈之时，如何能够有效地结合区块链技术，突破障碍闯出新天地，仍然需要在接下来的时间里不断努力。

区块链和大数据技术优化改造传统的供应链产业模式，还需要循序渐进、不断发展。虽然不会短期就脱胎换骨，但仍然前景可期！

6.7 数字资产的管理与交易

除了在金融领域的应用之外，区块链同样可以把各种实物资产、无形资产抽象成数字资产（Digital Assets）进行管理。

6.7.1 数字化实物资产的管理

不动产管理是实物资产管理中极为重要的一个领域。有数据统计，美国房地产业主每隔 5 ~ 7 年便会卖出他们拥有的房产，平均一生中会搬 11.7 次家，而每次交易都是费时费力的。更重要的是每次交易都需要一个中立的第三方参与其中，这个第三方的费用通常是房地产价格的 1% ~ 2%。而中国媒体报道显示，北京市房产交易中中介费大致为总房款的 2.7%，2016 年房屋中介费用超过 300 亿元。

国际比特币房地产协会（International Bitcoin Real Estate Association, IBREA）给出了一个取代政府土地及所有权登记的基于区块链技术的数字解决方案，包括以下三个方面：

- （1）房地产买卖；
- （2）房地产托管；
- （3）房地产所有权转移。

为了让用户更容易接受新技术，IBREA 的策略就是让用户在不知情的情况下使用基于区块链的技术。他们首先将房地产所有权记录都存放在区块链系统上，这个过程就像把文件上传到共享服务器一样简单。

用户在上传文件到区块链系统时需要证明自己的身份，而在复制该文件、进行所有权转让时需要证明他们现在是所有权人。可以说，IBREA 实际上就是借助区块链技术制造一个几乎不可能伪造的文档来防止欺诈。

当有人伪造了文件时，IBREA 的区块链系统将帮助证明你的身份。即使入侵者能够产生一个伪造文件，他们仍不具有转移资产的能力，除非他们也获取了你的专用私钥。因此，这里其实是有双重保险的：

- （1）现实所有权的实际安全文件；
- （2）可以转移财产的私有密钥。

除了不动产，贵重宝石也是区块链试水的领域。Everledger 是一个将区块链应用在钻石领域的项目，它将钻石验证信息和交易历史记录到数字化分类账本，可以有效防范钻石盗窃和保险诈骗；它通过创造包含钻石的切割、清晰度、颜色、克拉重量以及 14 种其他参照方式的数字 DNA，建立起透明的、不可篡改的钻石分类账本，包括其来源、物主以及所经历的过程。这个实践证明区块链对实物资产的数字化和证券化的适用性。

6.7.2 数据流通

京东万象

作为京东云倾力打造的大数据服务平台，京东万象目前已拥有超过 300 个数据提供商，超过 1 000 个数据源，也是国内率先应用区块链技术的大数据流通平台。

京东万象利用区块链技术搭建了联盟链，通过区块链数据不可篡改的特性把数据变成受保护的虚拟资产，确保每笔交易和数据都有确权证书。在数据进行交易之前，京东万象会先对卖方数据进行确权，明确其归属并将确权证书同步到各个节点上。而未经许可，盗卖则没有确权证书，或者证书与区块链确权不匹配的数据，数据提供方就可要求法律保护。这样既消除了数据提供方的担忧，也解决了数据需求方找到合规、正版数据的需求。

京东万象还要与公安等相关部门合作，建立个人数据授权体系，使个人数据可以在互联网上合法使用，以解决个人数据的授权问题。

Enigma

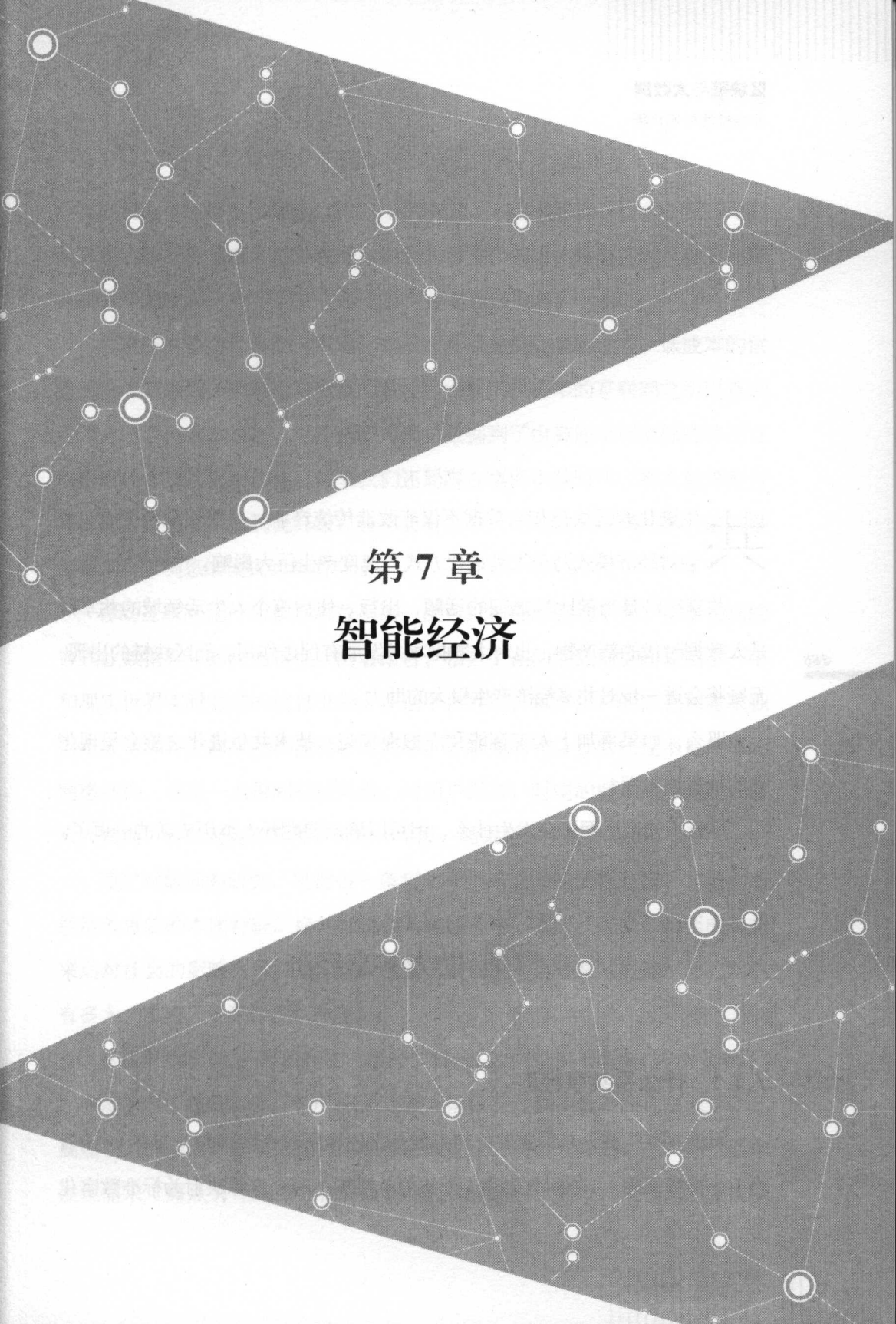
MIT 的区块链项目 Enigma 是通过在不披露隐私的情况下对数据进行分布式运算并输出结果的形式，解决了数据使用过程中的隐私问题。

大数据的交易则可以转变为对数据使用权的交易，数据产生时即以加密的方式被固定在区块链上。买方对数据的购买成为了触发针对特定数据计算的行为，计算的过程会消耗代币，而计算的结果则直接使用买方的公钥加密，



由买方持有。

数据在计算以及结果输出的每一步记录都会被留存在区块链上，不论是对数据源头的质疑，还是针对买方私自复制的追责，都可以通过使用区块链的可追溯特性来解决。



第7章 智能经济

共生进化的区块链和大数据不仅能改造传统产业，从更宏观角度看，也会对经济模式乃至人类进化方式与速度产生巨大影响。

共享经济是当前比较热门的话题，出行、住宿等个人生活领域的共享既是大数据时代的新产物，也对大数据本身发展有促进作用。而区块链的出现，无疑将会进一步对共享经济产生极大的助力。

那么，如果再加上人工智能和虚拟现实呢？技术共生进化之旅会呈现怎样的精彩景象呢？

尽管不知道后面还会发生什么，但可以确定要翻开人类历史新的一页了。

7.1 助力共享经济

7.1.1 什么是共享经济

所谓共享经济，其最重要的特点就是把原先所有权明确的（专有）、闲置的（等待被调用）、非标准化的（无法简单复制）零碎资源映射为标准数字化

的互联网信息来分享和整合,并充分调用起来。这和传统意义上的“借”不同,共享经济依赖的是对大范围数据和资源的掌控和调用。只有在大尺度上实现共享,才能成为一种“经济”,小范围只能是租借服务。

共享经济要在大尺度上实现,离不开两项关键的基础设施:低成本的信息流通(互联网)和对用户数据的掌控。发展了几十年的互联网之所以直到今天才孕育出共享经济,主要是因为用户数据到了由智能手机支撑的移动互联网时代才有了使用价值。如果我们还局限在桌面电脑时代,那么就算每台终端的接入速度都快到光纤级,也不会有几个人用电脑来叫网约车。这就是操作工具影响思维模式的典型特征。

移动互联网让行为和需求数据第一次可以精准定位到个人,常年在线的方式让数据有了即时的特性。两者结合,相当于在网络数据空间复制了一个和现实世界实时对应的信息世界。在这个世界里,每个用户的信息和需求独自组成个体特定需求所构建的映射网络。多维度的充分互联已经不是简单的网络结构,而是一点瞬间联通其他。对用户而言,时空的概念被黑盒化,被大大模糊,信息的应需而得只是操作手段,结果是服务应需而得。

通过对数据的研究,发掘每一条需求和供给之间的关系连接,才是共享经济最重要的本质特征。换句话说,共享经济能不能成长起来,乃至成长起来后对社会的影响有多深远,很大程度上就取决于这种联系能编织起的网络有多大、多密、多深、多少维度。

从这个角度看,目前流行的共享经济案例,如网约车(滴滴、神州专车等)和共享单车(摩拜、ofo等),只是这张大网编织过程中最初的几根绳子。其商业模式的建立,本质上和传统的互联网创业公司是一样的:在实际生活中发现需求(痛点),然后调用资源解决需求(痛点)。其中,依赖人力发掘联

系至关重要。这样的模式久经考验，但是强烈依赖创始人的商业敏感以及资金投入等因素。问题在于，这些因素无一不是稀缺的。那么，怎样降低共享经济的创新门槛呢？或者说，怎样才能发动更多的脑力参与到这张大网的编织过程中来，让编织的速度更快些呢？最简单的答案就是“人海战术”，让每个人都可以用极低的成本参与进来。

共享经济依赖的是对数据的调用。而传统模式下，数据本身被寡头垄断。且不说寡头们各自积累的数据之间由于重复造成的浪费，也不说全社会的数据被寡头们掌管后的安全问题，随着虚拟空间的进一步成长，其无孔不入的信息搜集终究会碰到大众对其界限的疑虑。所以，未来数据的存在和利用形式值得考虑。

那么，把数据作为一种全社会范围的公用资源，作为一种可以和水、电、公路相提并论的基础设施，就成为一种必然。供所有人使用的电力才能带来电气革命，供所有人使用的公路才能带来运输革命。同样，只有供所有人使用的数据才能带来数据革命。而共享经济就是数据革命的一种形式。

具体地说，这种作为社会基础设施的数据需要具有以下特征：

- (1) 合理的存储方式：高效的存储，便于任何人随时调用；
- (2) 可靠的保密特性：实现高度的保密性才能保证数据被所有人使用；
- (3) 合理的成本：低廉的价格是公共服务必须具备的特征；
- (4) 方便的可扩展性：随着海量数据蜂拥而至，数据系统必须能便捷地适应未来的新增需求。

7.1.2 共享经济与区块链

很显然，现有的数据组织形式根本无法满足上述需求。我们需要一种全新的方式来组织安排数据，这种技术就是区块链。

完全去中心化管理的比特币已经运营了很久，不论是数据的存储和调用方式，还是低廉的维持成本以及高度的保密性，区块链技术的稳定性都已经得到验证。只要设计合理，数据的可扩展性也可实现，更不用说可以随时新建更多的专用链来解决特种问题。由此看来，区块链技术完全可担大任，成为大数据的载体。

如前所述，从数据的角度看，共享经济的本质是在数据空间为需求和供给创建联系的服务。当下，个体的需求需要用户自己提出，个体提供的服务也需要用户明确量化。这无疑限制了共享的想象空间。如果数据成为一种公用产品，那么每个人都会时时刻刻产生海量的数据。而这些数据又可以在保留隐私的前提下汇集到区块链上，供任何有能力、有兴趣的企业或个人发掘。那么，借助对数据的挖掘，不论是通过人工的手段，还是通过时下火热的人工智能的手段，都有助于发掘潜在的机会。借用乔布斯的理念，既然用户自己都不见得知道自己想要什么，那就让大数据分析来帮助用户找到自己想要的。

191

共享经济之所以能兴起，主要原因有以下几个方面。

高效利用资源

共享经济是调用闲置资源，不依赖新的投入。从社会发展的角度看，其起到的是“开源”的效果。通过互联网对信息的低成本传递和高效处理，待售资源和服务就能高效地和随时随地产生的市场需求相匹配。从这个角度看，共享服务的平台方和证券交易所会有些许类似，提供的都是撮合交易的服务。从实际使用来说，因为使用的是闲置资源，所以前期投入更小，投资方（共享服务提供者）的回报压力也小，直接造成共享经济服务的成本低于传统企业提供方的成本。因此，当此低成本传导至用户时产生的是皆大欢喜的效果。例如打车市场，不论是通过自建还是和互联网公司合作的方式，这种即时匹

配司机和用户的做法在传统出租车行业并不难做到，但出租车队的固定成本则是无论如何也无法避免的。除了满足随时出现的需求之外，对周边环境的感知、市场趋势的即时预测、用户行为的理解都是大数据的拿手本领。就像电商可以用大数据提前调配仓储一样，大数据当然也可以用在其他应用场景。例如，提前预知打车需求的集中爆发，通过价格等方式调配司机。

信用系统

信用系统是应对非标准化服务的一大利器。如前所述，标准化的一项优势就是对质量的管控和保证。非标准化的情况下，为保证使用体验留住用户，最直接的方式就是建立信用系统。可靠的信用系统对提供方和使用方都是决定是否进行交易的低成本决策工具。需要说明的是，这一点的重要性随着交易金额和时长的增加而越发重要。和陌生人坐半小时的车相比，陌生人来家住一晚所涉及的考量就多得多了。可以想象，随着共享经济的深入，随着共享资源价值的变高，信用系统的重要性会越发加强，甚至成为成败的关键。

在现有共享经济的基础上，一旦数据区块链化，则直接实现的优势体现在以下几个方面。

政策和监管

传统模式的监管依赖的是威慑，靠的是对违规行为的探知和惩戒来完成对市场行为的规范。这就包含两大步骤：一个是对违规行为的探知，另一个就是探知到违规行为之后的搜证惩罚。因为两大步骤在传统模式下的时间成本和人力成本都很高昂，所以只能依赖威慑。实际上，对绝大多数违规行为都监管无力。而区块链技术则能提供电子化的、公开透明的、不可更改的记录，

大大降低违规行为的探知与搜证成本。实际上，基于区块链技术的智能合约技术可以强制实现参与方的商业行为合乎监管标准，大大提升违规行为的难度与成本。考虑到未来共享经济进一步发展，新型应用不断出现，高频、瞬时、海量的交易不断增加，传统监管模式必然力不从心，而区块链的特性能帮助监管部门方便在各个层面监控市场。

产业生态

区块链天然的分布式特性正是对抗中心化寡头的利器。原本必须依赖中心化寡头的交易撮合服务，完全可以交由基于区块链技术的公共交易系统来实现，彻底解放对专门平台的依赖。

与此同时，鉴于区块链可以低成本实现对信息的选择性分享，那么原本分散于各家的数据孤岛也可以自动打通。大数据在互联互通之余顺利地实现高度统一，对整个产业的发展 and 降低新兴服务的门槛都大有裨益。所以，自动进化的区块链技术平台必将取代大多数公司，人类社会的组织方式必将重新改写，这将包括全球范围的各种新型组织。

193

信用

在区块链透明的数据面前，伪造的难度大为提升。

(1) 解决支付担保的不便。

区块链与生俱来的支付功能用在共享支付领域，可谓是水到渠成。随着智能合约的进步，自动化支付更可以提升用户和商家的体验。

(2) 真实信誉评价。

有了基于区块链的可信赖的数据背书，无论是需求方还是提供方都可以

方便、放心地在预订阶段考察对方，促成交易。而记录在区块链上的透明和可追溯的记录，则可以方便地对抗虚假信用的产生。例如，通过在交易中添加电子签名、核实真实支付行为等都可以极大地增加虚假信誉的制造成本。辅以大数据分析，探知虚假信用的成本也能大为降低。

7.1.3 共享经济的发展方向

一经写入区块链，数据便自动实现“永生”。经年日久，区块链上沉淀的信息将是海量并可以直接使用的。当所有的行为都经由区块链记录，“历史”这个概念甚至都可以改变。在加密技术的帮助下，这份原始可信的记录在保证个人隐私的情况下供给所有人和企业使用，为互联网应用开拓了一个全新的阶段。最直观的商业信息包含各种交易记录、评价记录以及由此衍生的信用记录；次级直观的信息包括各种用户背景信息，如各类证件、医疗记录、交通违章记录等；再次级的信息则有一切个人行为数据。当这些信息得以公开，任何人都可以借用大数据的方式发掘其价值。区块链化的大数据也就变成了水电一样的社会公共资源，供共享经济调用。

最能受惠于共享数据好处的一类应用应该包含同一产业的上下游。例如，旅游业可以将食宿、交通以及各种附加服务整合在一起，而共享经济可以与其无缝衔接；餐饮业可以将物流、质量控制整合进共享模式，让家庭餐饮更加多样和个性化。甚至广告制作都可能因为共享经济而变得碎片化，专门服务于共享提供方的小额广告可能会出现。这种模式下，传统的标准化服务与个性化的共享经济可以取长补短，最优化用户体验。虚拟公司、虚拟产业、虚拟的融合产业必将大行其道。

当智能合约成熟之后，区块链会自然而然地从信息的记录者进化为交易

的参与者。低成本自动化的交易能极大地降低交易成本，结合物联网开发出之前想象不到的应用场景。当前的共享以汽车和住宿为代表，一部分原因是因为两者交易价值较高，承担得起相对较高的交易成本。当交易成本逼近0，而且实现自动化时，任何东西都可以随时随地以各种方式共享。辅以适当终端，甚至包括当前实现共享的工具——智能手机本身都可以实现共享。实现了智能手机的方便共享，也许就不再需要今天的大而全的智能手机，产品和服务的特色与差异化才是生存之道。

值得强调的一点是，区块链的公开特性和去中心化特性使全社会性的“大数据库”变得可行，这使真正的、不依赖于寡头的共享成为可能。届时，新兴服务可能会以碎片化的方式出现，依赖服务质量和特色野蛮成长，而非依赖寡头的平台提供的便利。这一点和音乐产业已经经历过的颇为类似。在线视频和用户生成内容的出现极大地影响了传统音乐电视，以往的巨星MV全球同步首发在今天已经看不到了，方兴未艾的网络视频和直播则说明了碎片化创新的活力。辅以区块链提供的便捷的知识产权保护和收费手段，碎片化的共享会更有活力。

195

7.1.4 共享经济的多种尝试

最先出现的惹火的共享经济项目就是 Uber 和 Airbnb，随后国内滴滴、神州专车迅速跟进，再加上其他公司一拥而上，网约车迅速成为了国内城市白领出行的重要选项。

短租房方面模仿跟进的国内公司也有很多，但像网约车公司那样成功做大的却没有。这一方面有中国国情的特殊性，例如，国内廉价连锁酒店太发达，出租方和承租方对共享观念的接受度不够，而最重要的是信任的缺失。

中国和欧美在社会信用上的不同有很多方面。但非常明显的是，希望通过政府部门和授权机构来授信，或者希望通过某些身处数据孤岛中的大企业来授信，这是不够的。最起码是无法提供足够的短租房交易中需要的双方授信。

当然，这里说的社会信用是指一般商家和个人可以方便、廉价地随时查阅的信用记录，可以在每次交易时被交易对象即时核实，而不是那种手续繁杂、调阅成本高昂（时间、精力也是成本）的束之高阁的信用记录。例如，政府授权的征信机构、各大银行和大企业自己独享的信用记录。没有真实、全面、不可篡改、全公开透明、方便查阅的包含社会个人和企业单位的全面社会信用，以共享经济为代表的很多商业模式是无法真正落地的。

最近火热的共享经济项目就是共享单车。和网约车这种“一手交钱、一手交货”的面清模式类似，共享单车对交易双方的信用要求不高，所以在没有方便的社会信用共享的情况下仍然可以跑起来。而共享汽车在中国如何有效落地，就需要在个人信用核查方面做出努力了。

接下来简单介绍一下其他多个行业的共享经济服务模式。

（1）宠物寄养服务：DogVacay 是一个提供宠物短期寄养看管的中介平台，解决用户出门旅游而家中宠物无人照看的问题。

（2）日常帮手服务：TaskRabbit 可以提供从整理衣物、帮忙购物、粉刷墙壁到组装家具等各种日常事务服务。

（3）家政服务：Zaarly 平台提供的服务有多种，包括从勤杂工到剪草坪等。每一个服务提供方都要接受平台方的面试，以及对客户的满意度调研才能登上平台。

（4）WiFi 共享服务：Fon 平台目前有来自全球 2 千万用户提供 WiFi 热点，作为回报，用户可以免费使用其他用户提供的 WiFi 信号。

7.1.5 展望

共享经济的核心在于调动闲置资源，从而降低全社会的成本。其本质是一种服务，而服务型经济则是目前所有工业化国家发展的方向。服务的目的是改进原有进程，润滑经济运行，改进用户体验。现有的服务型经济在很多时候背离了这一点，以自身利润最大化为终极目标，偏离社会需求。当下正在进行的从传统服务器到云计算的转变过程就能体现这一点。对于全社会来说，这并非最优解。

共享经济如果止步于用中心化的新寡头代替原有寡头，则不会给整个社会带来根本性的变化。事实上，鉴于其规模和运作方式，很可能会造就一批更难被革命的寡头。所以，为了让共享经济充分体现出其潜力，在各行各业全面开花，打破数据的藩篱是一种必然。

197

传统模式下，用户对个人的敏感信息上传是疑虑重重的，更别说主动制造数据隔离的大数据企业。而区块链的安全性和匿名性则可以最终解决这种顾虑。区块链的公开性和去中心化特点则具有从根本上实现全社会范围内的“我为人人，人人为我”，在不增加投入的情况下真正让每个人都体会到资源优化配置带来的便捷与廉价，降低社会运作成本。

7.2 插上虚拟现实的翅膀

7.2.1 又一个热点技术：虚拟现实

这里我们又迎来了另一个高德纳热点技术——虚拟现实。在第1.3.1节中

的技术成熟度曲线上，虚拟现实技术已经到达了幻灭期的谷底附近，开始了真正走向产业化和商业化的步伐。

广义的虚拟现实包括通常意义上的虚拟现实（Virtual Reality，VR），也包括增强现实（Augmented Reality，AR）和混合现实（Mix Reality，MR）。

首先，虚拟现实可以简单看作是一种界面技术。人类与外部世界交互的直接手段是眼、耳、鼻、舌、身、意，眼镜、望远镜、电视、电脑、手机等都是逐步发展出来的交互方法，也是信息综合在一起与人交互的界面。所谓“秀才不出门，全知天下事”，当代“秀才”只要掏出手机来就可以了。

其次，虚拟现实技术和可穿戴设备、传感器技术有很大的重叠。随着科学计算和工业制造的发展，人类对内、外部世界的认知有了更加丰富的手段。

X光、红外线、核磁共振、价格低廉的高清摄像头等新设备，配合深度学习自动辨识图像的能力、眼动追踪有效获取关注焦点的能力等，这一切大大拓展了人类身体的能力，增加了可获得数据的丰富性，实现了无论量级还是维度的全面提升。而且，只有通过虚拟现实，一个真实世界的完备映像才能真正形成，数字化虚拟世界的融合与进化速度必然远超物理世界。

最后，虚拟现实不仅是单向的信息传递，还可以提升人类影响内、外部世界的能力。战斗机飞行员的数字头盔就是典型的案例，所有信息在眼前随需提供，战斗的感觉越来越像打游戏了。而区块链技术加入其中，会使军事数据链路更加可靠。

7.2.2 结合了区块链和大数据的虚拟现实

在戒毒治疗中，通过递进方式播放虚拟现实影片，让吸毒者对毒品的依赖转化为对毒品的厌恶和恐惧。整个治疗过程中，除个别人因为自身的身体

条件在观看 VR 影片时会产生轻微的眩晕感之外，并无其他副作用。

从心理治疗角度而言，听觉比视觉更有激发共情的可能性。因此，在短片的制作中着重加入了许多对吸毒人员有固定含义的环境音、对话等，以提升治疗矫正效果。

类似的临床实验案例显示，60 余名吸毒人员在经过 15 天的 6 次虚拟现实系统治疗之后，对毒品渴求度降低的比例达到 75%；而同一时间内，未使用虚拟现实系统治疗的对照组人员仅有 3% 降低了对毒品的渴求度。

国家禁毒委员会、国家禁毒办相关数据显示，近年来，我国滥用合成毒品的人数急剧增加，年均增长达 36%，2015 年新发现吸毒人员中 80% 是滥用合成毒品人员。业内人士认为，从这套 VR 戒毒系统的仪器反映的数据看，该系统在矫正“心瘾”方面有一定效果，但仍需要更长时间的验证。

199

戒毒需经历生理脱毒、教育适应、康复巩固、回归指导四个时期，VR 戒毒系统是科学戒毒中的一部分。浙江省相关机构依托该系统搜集的数据建立起千人级别的毒瘾矫治大数据库，通过对大数据的不断采集与积累，为后续的升级研发和治疗提供了数据基础，以达到精准戒治的目标。

而戒毒是隐私性非常强的应用，信息必须真实可靠，但又需要匿名，这是最常见的应用场景，也恰恰是区块链可以大显身手的地方。区块链保护隐私而又消除数据孤岛的巨大优势，为大数据在戒毒等领域的应用奠定了坚实的基础。

和戒毒非常类似的是心理治疗，无论是真正的精神卫生方面的问题，还是心情不好需要排解忧思，虚拟现实都需要在大数据和区块链的帮助下，一方面充分保护个人隐私，真正打开求助者的心扉；另一方面需要依赖大数据卓越的发掘、辨识和分析能力，集全社会的成果和优势，真正服务和帮助到每一个人。

7.2.3 可视化沉浸让思维开始跳跃

无论多么复杂完备的数据和精良的处理结果，都要以人为本，归于人文主义，最终还是要人来思考理解和创造开拓。虚拟现实是所有技术的黏合剂、融合剂，把包括区块链和大数据这几种技术加以融合，才使整个当代科技的立体网络结构开始完备起来。

技术没有对错，没有高低贵贱，只有是否恰当和适配与否。单独谈论某项黑科技往往是要流氓行为，踏踏实实地综合解决方案、帮助解决不同应用场景下的问题才是王道。当然，还不仅是解决当前问题，创造新的应用场景更是自然演化的必然结果。

心理学研究早就证明，人类感知学习把握内、外部世界，如果可以把全身感官统统调动起来，远比常见的用脑记、用眼看、用耳听的效率要高得多。这一点，虚拟现实有得天独厚的优势，3I（Immersion：沉浸感；Interaction：交互；Imagination：构想）本来就是突出的特征。

“我思考问题时，不是语言，而是用活动跳跃的形象。”类似的话，爱因斯坦和特斯拉都说过。虚拟现实可以改变的是人类整体的思维模式，把少数天才的优势变成普通人的常态。用虚拟现实帮助我们以真实世界的图景和轻松调整变化的立体模型与场景来思考问题，这已经是业内共识。

但这里问题的核心是我们获取数据的能力快速进步，而呈现方式却一直是比较原始的二维平面方式，如常见的 Excel 表格、条形图及饼图。虚拟现实允许我们“在数据场景中行走”，快速变换任意角度观察数据之间的联系，这本质上是打通左右脑，回归人类对图像的超强处理能力。

在这方面，英国 Masters of Pie 和 3D 软件开发咨询公司 Lumacode 进行

了非常有益的探索，合作赢得了 2015 年大数据虚拟现实挑战赛冠军。他们的获奖产品让用户可以通过虚拟现实头戴式设备操作大数据系统。他们的研究人员甚至可以在虚拟环境中对最小的几何模型进行抓取、移动和点击，提升对数据交互的控制能力。

7.2.4 虚拟现实高效连接数据与人

用虚拟现实更有效连接数据与人，小公司在这方面的创新尝试很多，大公司更是不甘人后。谷歌有数字地球的雄厚数据基础，结合虚拟现实技术，并经过谷歌眼镜的尝试后，目前重点在发展 Daydream。

移动时代落后的微软力度更大，HoloLens 增强现实眼镜是业界真正的标杆产品，虽然价格三千美元，但是实用性已经很强。蒂森克虏伯电梯公司在其博客宣布，微软和他们达成合作，为 24000 名技术人员配备 HoloLens。对于黑科技产品来说，这种单一销售量还是很能说明问题了。另外，HoloLens 在医疗等行业的应用也有很多典型案例。

微软的重心在于商业：“Transform your business”。除了帮助一些勇于创新的企业转型，如沃尔沃、Trimble、奥迪、NASA 等，微软更直接去攻坚每个行业的龙头企业，给 Hololens 树立了一个个极具代表性的案例。

全球最大的骨科及医疗科技公司之一——史塞克（Stryker）成立于 1941 年，比微软还要早 43 年。Hololens 应用于医疗培训、健康检查、辅助脑部手术，这些已经为人熟知，现在又增加了一项崭新的应用——史塞克用 Hololens 设计手术室。一般而言，泌尿科、骨科、心脏、耳鼻喉科和普通的手术都会共享手术室，导致配置非常复杂。传统的手术室设计要全员出动开会，频繁交互，低效费时。有了 Hololens，外科医生可以直接在办公室的桌面上讨论设计，

而工作人员可以根据医生的设计随时修改模型。

在设计行业依赖虚拟现实，相比依赖想象力的平面图纸自然强大很多，启动混合现实功能把自己放进场景里面，就是进入全新世界的感觉了。词不达意、言不及义的种种沟通错误会减少很多，让沟通成本大降，效率显著提升。那么，人机工程也就不再依赖想象和烦琐的实体建模了。

7.3 拥抱人工智能

写作本书的时候恰逢 DeepMind 阿尔法狗（AlphaGo）的升级版 Master，利用人工智能算法在网络上挑战众多世界级围棋大师而大获全胜，这让人深深感觉到人工智能的山雨欲来之势。一系列的新科技和新技术将给人类的生活方式、生产力发展乃至社会治理模式带来翻天覆地的变化，而人类就像在功能手机时代那样，根本无法想象几年后智能手机可以给生活方式带来的巨大冲击。

随着人机界限被突破，人工智能不仅是一个工具，更是跨物种合作的一扇门。区块链、大数据与人工智能的结合，将使我们从生物生命体进入人类生命共同体，最终进入超生命共同体或智慧生命体。

7.3.1 大数据孕育了人工智能

人工智能涉及的领域非常广泛，并且已经深入人们的生活。基于大数据的积累和应用，人们开始在其中发现某种规律，引发了分析的需求，让机器开始有了思想。当硬件性能逐渐提升、计算资源越来越强大时，大数据孕育

了今天的人工智能。很多过去只有人能做的事情，现在逐渐能够通过机器实现，典型的案例包括语音助手、无人驾驶、机器人。

AI的基础是大数据，但这些资源通常掌握在巨头手中。这也是为什么这个领域的头条新闻总是被微软、谷歌、IBM、苹果、亚马逊、Facebook 这些公司抢去的原因。在国内，BAT、京东这样的企业同样拥有足够的用户基础，并且已经开展了应用。

人工智能本身就是大数据进一步发展必须结合的伙伴。大数据采集的不仅是规范化的数据，还有大量非结构化的数据和不同维度的庞杂信息。

结构化数据可以存储于普通的数据库之中，相应的处理会比较规范和简单。而非结构化数据涉及各种信息来源，在大数据时代占比会急剧上升。特别是随着社交媒体的兴起，非结构化数据更是迎来了爆发式增长。真正要对这些大数据进行分析并不简单，无论是自然语言处理技术，还是图像解析技术或者语音识别技术，都是传统意义上人工智能研究的范畴。没有人工智能的帮助，大数据很难做高水平的处理和分析。伴随阿尔法狗被人们所知的神经网络深度学习，是人工智能的更进一步发展。人工智能不仅需要对已有信息的分析解读，更需要建立主动的信息获取和学习能力。

显然，人工智能和大数据是相伴相生、共生发展的又一个典范。随着更多数据的积累和更好硬件水平的支持，人工智能将会迸发出更强大的活力。

7.3.2 让人工智能的价值真正流转起来

区块链网络传递的是信任和价值，而价值是无处不在的。人工智能可以被看作是传统大数据资源的高级阶段，数据已经不再是一个个初级商品（如

矿石、原材料、能源)，而是开始因融合而产生更高的附加价值（如变成了机器设备和生活用品）。显然，人工职能所需要的价值体系和价值交换网络将更复杂。

价值网络信誉

无论是人类成员，还是人工智能或机器设备，区块链可以为网络上的所有参与者分配特有的身份。这个身份不仅是一个 ID，还有背后的信誉，包括传统意义上的信用和非信用范畴的口碑、认知等。本质上，在区块链网络中被交易的是各个 ID 信誉背书的各种价值，没有信誉就没有价值。与其说区块链构筑的是一个价值互联网，不如说构筑的是价值社会的行为规则，而且是超越了人类范畴的价值社会，也包括人工智能。

随着人工智能和虚拟现实等技术的发展，人类在升级版虚拟社区的数字化足迹将成为人类行为和情绪（甚至表情、脑电波等）方面的大数据的一部分，进而成为价值网络信誉的一部分。以区块链、大数据、人工智能、虚拟现实等技术为基础的 360 度全维度价值网络信誉的建立，可以颠覆性地改变人类社会的运行。

价值交易配置社会资源

在价值网络社会的各种经济活动（交易）是社会资源配置和社会权益保障的手段。这些可以交易的资源，不但包括现代人类社会的各种有价资源，还包括现在无法定价、无法交易或者不适合交易的大量不可交易资源，如人类的闲暇精力、各种非商品化的个人物品。而通过区块链智能合约，还可以方便地实现当前交易成本极高的复杂交易模式。

由各种交易引导的价值导向，让社会资源可以更社会化、更敏捷、更可实现地优化配置。本质上，现在的很多政府管理职能是由于社会化自发运行和管理无法实现、强行插入的外部干预。如果有社会化运行方式，确实可以省掉诸多政府管理任务。

无论是资源的敏捷配置，还是交易的智能中介，更包括对新资源价值发掘和新经济模式的开拓，大数据和人工智能都可以起到巨大的作用。大数据不再停留在机械的统计功能和分析功能，还包含在人工智能的帮助下主动参与引导、主动介入社会活动，从而创造更多有价值的行为。因此，智能交易将成为社会管理优化发展的重要步骤。

有信誉背书的社会运行

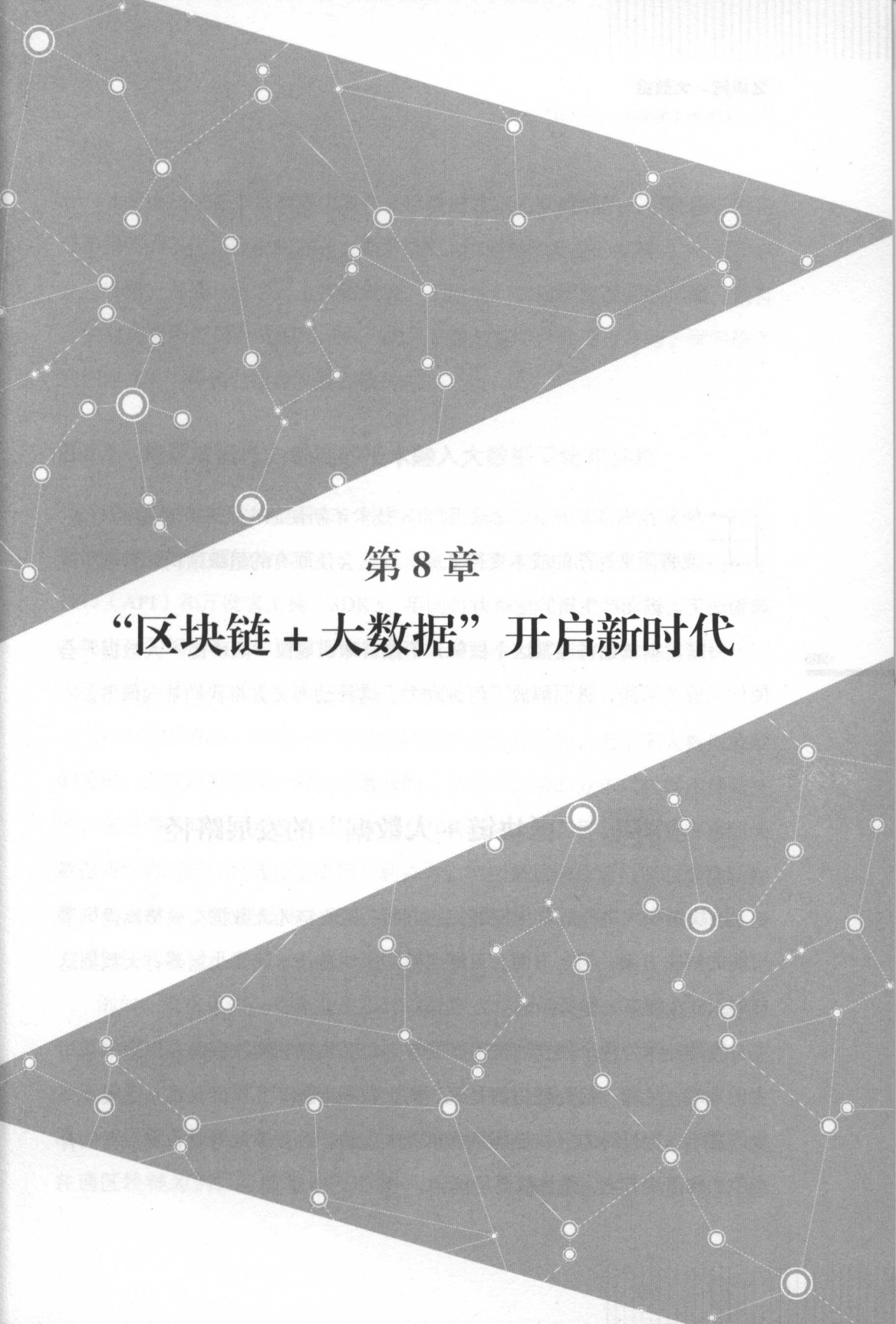
在此基础上形成的新的社会化管理规则，就将一举超越现今区块链领域冷冰冰的智能合约，也会超越今天的社会公德范畴。

不可篡改的社会信誉所背书的价值交换会为交换的价值赋予信誉权重，而这会投射到每一笔交易和每个个体（人类或者机器）的社会生活的方方面面。当今社会，人们对社会公德的不在意是因为很难由此导致利益受损。法律监管的高昂成本导致其在大多数社会活动中只起到底线作用，而无法参与调节。区块链网络筛选的不仅是人类的信誉，还可以包括设备和人工智能。恶劣的人工智能会被网络淘汰，也许这也可能会投射到技术进步的方向。

有信誉背书、有价值权重来驱动的社会运行是大数据和人工智能大显身手的舞台。也许那个时候我们要说，是否有价值已经无关紧要（每个事物都有价值），但是否能真正发现在特定场景下的价值，是否能为事物在

特定场景下赋予更多的价值，这些才是关键。而这正是大数据和人工智能最擅长的。

当前，我们还处在区块链技术、人工智能和大数据逐步发展的早期，很难估计未来这些技术的融合会带来什么。但有一点可以肯定，未来的这种变化一定远远超出我们今天的想象。



第8章

“区块链+大数据”开启新时代

任何新技术改造社会，都是因为其技术革新使原来的不可能成为可能，或将原来高昂的成本变得低廉。而这会使原有的组织结构和管制方式成为历史，进而产生新的生命力。

208

所以，本章也将遵循这个框架，不揣冒昧，妄议“区块链+大数据”会使什么成为可能，进而解放了何种动力，这种动力又会将我们带向何方。

8.1 “区块链+大数据”的发展路径

区块链和大数据的共生发展，一方面，区块链为大数据突破樊篱提供了可能的解决方案；另一方面，日渐成熟的区块链技术确实也需要有大数据这样的大江大海来一显身手。

但是，作为两个独立的技术发展方向，区块链和大数据的合作也不是与生俱来的。忙着“捡到篮里都是菜”的大数据折腾了几年，存在系统版本升级的需求。而投标方区块链能不能承受大容量、高频率的考验，真正提供有竞争力的落地方案，是挑战更是机会。

下面谈到的五个层次可以看作是区块链技术走向成熟的五个阶段，每一步所需要的不仅是商业的推进，更需要技术的成熟和社会的发展。

当然，人类一思考，上帝就发笑。无论我们如何严密逻辑的预测，也有可能被历史所证伪。我们可以做到的只是参与其中，并通过思考、学习给予我们自身的宁静和对社会发展的敬畏。

8.1.1 将区块链作为单纯的技术融入大数据采集和共享

如果简单地将区块链看作一种分布式数据库存储技术，其实就是一种底层技术支持的数据结构和接口，并提供一套与开发语言无关的标准应用程序接口（API）和开发者工具（SDK）。不同时间、不同技术和不同语言开发的各类应用和相应的操作型数据库，都可以通过不太复杂的步骤将重要信息写入区块链，并可以从区块链上获取已有的信息。

打破数据孤岛，形成一个开放的数据共享生态系统，是未来大数据成败的关键。而区块链作为一种不可篡改的、全历史记录分布式数据库存储技术，在强调透明性、安全性的场景下自有其用武之地，可以有效解决当前大数据遇到的问题。而这会驱使相关利益方，特别是政府或者行业联盟推动打破相关利益者的数据孤岛，形成关键信息的完整、可追溯、不可篡改并多方可信任的数据历史。

例如，很常见的一些企业违反环保规定，在夜间关掉环保设备而偷排，希望通过违法的方式减少生产成本，获得更多的经济利益。这种情况举证并不容易，虽然布置了很多监控传感器，但传感器的数据可能会被事后修改。如果所有环保设备和相应监控传感器的数据都被实时写入区块链（以井通区块链为例，目前每6秒形成一个区块），则这些数据无法被篡改，

并且被分布式地存储在全网不同的节点上。这些节点除了厂家之外，还可以部署在监管部门和能被公众实时查询的场所，由此产生的监管效果必然极为显著。

通过将区块链作为一种分布式存储的统一数据结构和接口，可以用比较低的成本来实现关键重要数据的互联和共享，一定程度上打破数据孤岛并形成多方信任的数据链条。

8.1.2 将区块链作为数据源接入大数据分析平台

区块链的可追溯性使数据从采集、整理、交易、流通以及计算分析的每一步记录都被留存，使数据的质量获得前所未有的强信任背书。这保证了数据分析结果的正确性和数据挖掘的有效性。

因此，从区块链上获取数据作为大数据分析的补充是应有之义。尽管大数据的发展趋势使对大部分类型数据的精确性要求降低，但是对于某些追求正确性的重要数据，把不可篡改的区块链作为数据源就很有必要。

最近落地的互联网金融监管新措施，要求监管同一单位或个人在所有互联网金融平台上的融资上限。如能通过（监管部门）统一的区块链网络（见图 8-1），把同一主体在所有互联网金融平台的贷款余额作为重要数据加以记录，并和其他大数据信息一起分析，监管部门就可以有效地进行监管，各类商业机构也可以有效控制自己的风险。

数据隐私保护一直是大数据发展的一个掣肘，大数据时代所需要的数据互通、数据共享实际上和保护个人隐私之间是存在剧烈冲突的。区块链技术通过多签名私钥、加密技术、安全多方计算技术，就能够只让那些获得授权的人才可以对数据进行访问。数据统一存储在去中心化的区块链或者依靠区

区块链技术搭建的相关平台，在不访问原始数据的情况下进行数据分析，既可以对数据的私密性进行保护，又可以安全地提供社会共享。

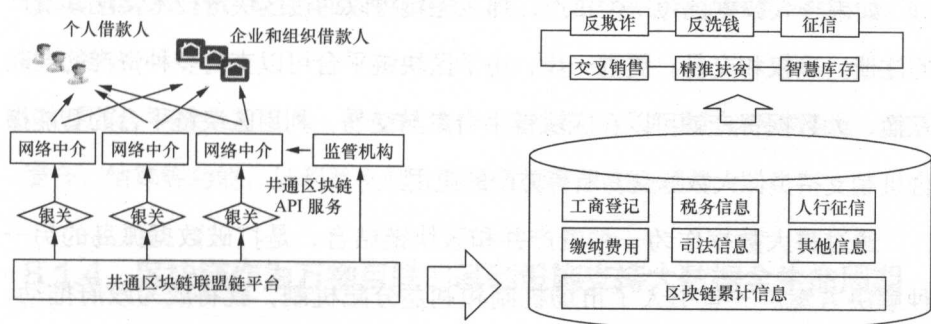


图 8-1 基于区块链的金融监管平台

数据隐私保护一直是大数据发展的一个掣肘，大数据时代所需要的数据互通、数据共享实际上和保护个人隐私之间是存在剧烈冲突的。区块链技术通过多签名私钥、加密技术、安全多方计算技术，就能够只让那些获得授权的人才可以对数据进行访问。数据统一存储在去中心化的区块链或者依靠区块链技术搭建的相关平台，在不访问原始数据的情况下进行数据分析，既可以对数据的私密性进行保护，又可以安全地提供社会共享。

因此，将区块链保存的数据作为大数据分析的数据源，为大数据分析补充精确的关键数据，同时利用区块链的匿名特性在一定程度上保证数据隐私，可以为大数据的发展提供关键性的帮助。

8.1.3 将数据作为一种资产在区块链网络中进行交易

区块链是从比特币开始为人所知的。可以说，比特币是区块链技术第一

个成功的杀手级应用。在目前的第三代区块链网络上，可以将任何资产数字化后进行注册、确权、交易，智能资产的所有权是被持有私钥的人所掌握，所有者能够通过转移私钥或者资产给另一方来完成出售资产行为。

如果将大数据视为一种资产，那么无疑可以通过区块链技术实现其资产的注册、确权和交易。更进一步，由于区块链平台可以支持多种资产的互联互通，大数据资产就可以在区块链平台参与交易，利用区块链平台的智能撮合机制支持类似大数据交易所等方面的应用。

通过将大数据作为一种资产并和区块链结合，是打破数据孤岛的另一
种解决方案。一旦引入了市场机制和利益分配机制，就将成为政府推动、
行业自律以外的另一种强有力的动力，促进大数据的实质性流通和产业化
应用。

212

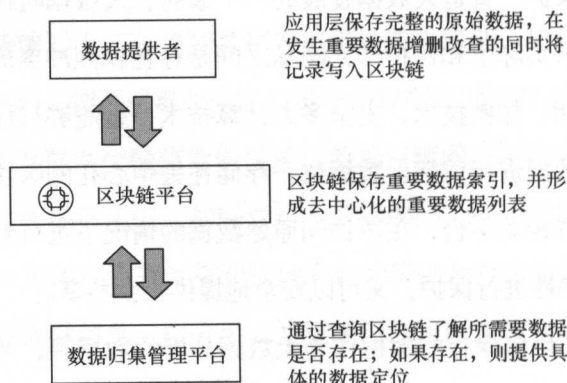


图 8-2 区块链存储数据机制

图 8-2 展示了利用区块链存储数据的机制。利用区块链不可篡改的特性，在区块链上形成具有时间戳的重要数据存取读写痕迹，保证数据归集平台可以获得链上所有重要数据的存取记录，从而定位需要查询的数据；对于特别

重要的数据，也可以直接保存在区块链上。

在此基础之上，还可以通过在区块链上部署智能合约的方式实现与数据共享相关的业务处理逻辑，包括租借、购买、转账等。智能合约在完成逻辑编写后自动进行，实现对数据共享的全自动交易和后续的账务服务，如发放积分等。

所以，将大数据视为一种智能资产，区块链可以提供包括资产注册、交易、撮合、结算等功能，从利益机制上促进数据开放和应用开发。

8.1.4 区块链作为万物互联的基础设施支持大数据全生命周期

区块链作为一个去中心化的网络平台，可以包含全社会各类资产，让不同的交易主体和不同类别的资源有了跨界交易的可能性。在这个价值网络中，不但可以进行传统的商业活动，还可以进行非商业的资源分享。只要你在该网络平台上具有信用，只要你的资源有价值。而区块链技术可以保证资金和信息的安全，并通过互信和价值转移体系达成了此前无法完成的各种交易和合作。

我们相信，未来区块链将类似如今的互联网成为价值互享的基础设施，人类的大部分经济活动可以用终端用户不易察觉的方式运行在区块链上。行业的上下游通过区块链共享供应链信息并进行智能生产；各类资产在以区块链为底层技术的交易所进行交易和互换；陌生的多方可以基于区块链上的可信记录进行合作；政府的公益和社会慈善通过区块链增加公信力和透明度。

在此基础上，区块链既成为各类经济活动的基础设施，同时也是各类数据产生的源头。区块链从技术层面不仅可以提供不易篡改的数据，同时也提

供了不同来源、不同角度和维度的数据。大数据分析可以基于全网的分布式存储的结构化数据和非结构化数据（通过哈希保存在区块链上），并通过新的存储技术（如图像化存储）增大容量。

当这一天到来时，区块链可以为共同的价值互联网提供高质量、经过稽核和审计的数据，而区块链本身则从大数据分析的补充数据源提升为大数据生命周期的主要数据源。

8.1.5 智能合约和大数据促进社会共治

随着数字经济时代的发展，大数据能够处理越来越多的现实预测任务。而区块链技术能够通过智能合约，通过 DAO、DAC、DAS 来自动运行大量的任务，帮助把这些预测落实为行动。

未来社会治理中，地方政府作为资源供给方，在进行诸如精准扶贫、社会服务外包、公益管理、养老等方面，都可以通过区块链作为中介，通过大数据作为公共产品需求者的精准分析工具，通过智能合约为标准化的公共产品提供自动流程。这不但可以大大减轻地方政府所需编制，也可以优化和提高社会服务的水平。

在这个条件下，有志于公共服务的个人和团体与有需要的人群对接，政府则通过区块链将相应需求内容和服务情况记录下来，并给服务方提供相应的报酬。而这一切可以做到透明公正和可审计，通过智能合约可以减少人为干预和冗长的审批环节。

以某公益积分项目为例（见图 8-3），大数据分析精确定位和归因不同的贫困户及需要帮扶的群众后，一方面，通过建设区块链上的公益积分体系，



8.2 “区块链 + 大数据”实现新的社会融合

8.2.1 对旧体制的反思和对新方向的探索

对旧体制的反思

区块链源于社会大众对金融体系的信任危机，因此当时奉行的是极端的去中心化模式。但人类社会的结构则是发源于人类成员之间的相互信任。所以，正确的做法是从技术上解决系统性风险，同时在法律、制度上加强监管，而不是彻底放弃中心化，完全丢弃人与人之间互信这个社会基石。

不论是人类社会、动物群体，还是全球大气现象，都更近似于复杂网络的架构，都体现出非标度网络的特性。这种网络中的各节点位置和彼此的交互关系都是唯一性的。这个特性也决定了当区块链技术与人类社会融合时，不能只通过完全去中心化的单一模式。

不论社会伦理，还是系统特征或技术发展，都要服务于社会需要、促进社会融合这个大原则。区块链与大数据相结合，真正应该达到的社会目的也是敦促社会进行改革、增强自我约束、重新建立社会的信任，而不是彻底丢弃人类的信任来开发出一个完全无信任感的技术系统。

对新方向的探索

大数据是继传统 IT 之后下一个提高生产率的技术前沿。在人类希望持续

地从数据中获得洞察力的推动下，大数据需求被不断提出。从这一点来说，大数据是人类经验和行为历史记录的精确定化和高频化，其背后的动力就是人类依然希望从过去预测未来。

大数据改变了人类可能获取数据的方式、频率和容量，大大增加了人类了解现实的能力，因此也改变了人类使用历史经验预测未来的模式。在这方面，笔者推荐维克托·迈尔-舍恩伯格的《大数据时代》对改变所做的概括。

(1) 从采样样本集到全量数据收集。

在大数据之前，人类不得不依靠随机采样的方式建立样本集，并在其上各种数据建模和分析。而如何采样保证样本集可以有效代表全量数据，就成为统计的一个重要主题。单单如何生成真的随机数就有大量文献可查，任何一种采样都会由于算法和模型各自的问题而存在模型风险。大数据使不进行采样而就某一个单一问题获得全量数据成为可能。

(2) 从追求绝对精确到降低精确需求。

由于巨量数据的存在和数据不断更新频率的加快，对所有数据获得完全精确的结果会成本巨大而收效不高。相反，相对不那么精确的数据也可以用来提供完整的趋势分析。同时，降低对数据正确性的要求使可以获得的非结构化数据大幅度增加，其背后的逻辑是大数据的简单算法比小数据的复杂算法重要。

(3) 从理解因果到分析相关。

由于大数据分析可以获得的预测大部分以两件事情发生的相关性表示，又由于大部分场景下了解相关性已经足够获得相应的业务收益，所以过去我们期望从数据分析中得出因果关系，在大数据情境下则只需关注相关关系。

基于以上三点，可以发现所有的变化都是来自技术的更新以及新技术下的成本关系。由于数据采集技术的提高，全量数据的获取已经成为可能。相

应地，运用全量数据并使用现有简单算法分析的效果也足够好。因此，新的复杂算法的开发成本就可以节约了，相同的投资不如去建立全量数据集。

而更有争议的应该是第三点，即相关关系取代因果关系。可以看到，这种趋势的存在也是由于大数据下相关关系的获取成本更低，通过不断提高的计算能力可以快速捕捉上万个指标的相关性，并支持从超市布局到新书推荐的商业决策，同时带来收益（交叉销售成功率提高）和成本缩减（原来从事这些推荐的专家团队解散）。

如果赌注够大，例如判断一个人是否犯罪或者决定一笔巨大的投资，是否仅仅依靠相关性就足够呢？因此，针对足够重要的问题，因果关系的分析依然重要，而机器识别的相关性将是分析因果关系的重要参考（相关性的解释在数学上可以使用多因子模型等方式，即两个变量拥有相关性，其实质可能是巧合，也可能是由于背后有相同的驱动因子——后者就是一种因果关系）。

如何看待经验

我们应该认识到，大数据分析的本质依然是通过历史和过去分析、预测未来。人类的这个习惯一直没有改变，但是我们的经验在多大程度上是可以依靠的？即使经过了大数据分析之后。

一方面，大数据无疑提供了更多维度和更全面的历史记录，以及在此基础上利用算法可以获得的各类分析结果；另一方面，大数据的预测，尤其是相关性预测更多提供的是一种基于概率的分布，而不是确定结论。这不是大数据分析的错误，但是使用者要心里明白。

电影《赢在球生》展示了数据比经验更准确的一个场景，那就是当棒球队不再通过所谓专家，而是通过分析申请人的各类细化指标来选择球员，从

而大大增加了选择到运动健将的可能性后，就让那些“专家”理所当然地丢掉了饭碗。

另一部电影《少数派报告》则讲述了当大数据分析（电影里是三个具备超能力的人）揭示某人将要但尚未犯罪时，我们在多大程度上可以依靠这个预测来定罪？你能因为某人有80%的犯罪可能性而给他定罪吗？那么90%或者99.9%呢？

8.2.2 新一轮的社会改良和变革

纵观历史，划时代的技术创新总会带来颠覆性的社会变革。但是，“区块链+大数据”技术有望改变这个传统，以一种和平的方式实现新旧时代的平滑过渡。

商业模式

219

区块链技术与现实社会的结合不同于从无到有新建网络世界，而是基于已有架构、软硬件布局的低成本耦合。

因此，区块链的使用并不会颠覆现有的生产要素构成，也无需改造现有的互联网基础结构。区块链是通过盘活存量，发掘现有条件无法直接利用的网络资源，塑造出的一种全新的社会组织形态和商业模式。

分布式

当中心化的云计算或大数据方兴未艾时，作为其对立面的去中心化的区块链也风起云涌，正好是中国人阴阳平衡思想的体现。

因此，简单地说，区块链和大数据两者是伴生和互补的关系，互为共生发展，也就是所说的有效去中心化。

这种模式既能够享受去中心化的安全和成本优势，又不至于过度地为了去中心化而降低效率——通过技术与现实的融合实现了正和（而非零和）游戏。

技术优势

区块链提出了不依赖第三方的技术信用，为重塑社会信用体系提供了解决方案，建立了独特的技术优势。

另外，区块链技术具备身份公开与匿名并行、数据开放与封闭共存等特点，可以扩展出无穷尽的、符合社会习惯和商业原则的应用场景，从而更有效地促进时代融合与技术革新。

社会治理的自然需要

美国白宫发布了 2014 年全球大数据白皮书《大数据：抓住机遇、守护价值》，报告鼓励使用数据以推动社会进步。

如果能够富有创造性而有效利用大数据来提高效率和质量，预计美国医疗行业每年通过数据获得的潜在价值可超过 3000 亿美元，能够使美国医疗卫生支出降低超过 8%；利用大数据实现政府行政管理方面的运作效率提高，估计欧洲发达经济体可以节省开支超过 1000 亿欧元（这其中尚不包括可以用来减少欺诈、错误以及税差的影响作用）。

8.2.3 孕育新的经济发展方式

基于云端的三流合一

2016 年 9 月召开的杭州 G20 峰会提出：“通过数字经济等新方式，为世

界经济开辟新道路，拓展新边界。”新兴的区块链技术和大数据结合，可以完美地契合数字经济的发展趋势。

第一，“区块链+大数据”可以低成本地实现信息流、资金流和物流的三流合一，从而有效降低社会的运营成本。

第二，技术接口的简洁性、业务模型的通用性，使其具备无限广阔的使用范围和扩展性，可以方便快捷地涵盖社会发展的各个方面。

第三，数字经济的核心是数字资产定价，数字资产合法化的核心则是权利方（资金流）转移。“区块链+大数据”可以用技术保证数据互信，实现交易背书，为最终构建数字经济世界的资产交换和价值交易提供技术保障。

可以大胆地设想一下，未来的某一天基于区块链技术的价值互联网将作为生产关系的一部分，对社会生产力产生重大的影响。因此，从宏观意义上讲，区块链技术是一种可能会对人类社会产生重大变革的技术。区块链技术对数据的可加密和可公开，用户身份的可匿名和可实名，通过排列组合可以实现无穷尽的场景。这给予了大数据更多维度的便利性和灵活性，可以满足不同的商业需求。

面对如今大数据所带来的巨大商机，引入区块链技术作为依托，就有可能突破当前大数据发展的一些痼疾，有机会在商业竞争中脱颖而出。

被重新定义的社会个体

第一，在当今的互联网世界，每个人（节点）都是独一无二的，这是大数据时代来临的重要标志。

在前大数据时代，由于信息维度的单一和极为有限的信息交互，互联网的匿名性得到了很好的体现。而在今天，由移动互联、社交网络、电商支付、

无纸办公连接的每个人有了越来越多的维度和周围的人进行交互，越来越清晰的个人画像将会呈现在世界上，匿名已然不可能。

这就意味着每个人更多的个性化资源和独特价值将会在大数据时代被有效发掘，此前由于个体信息量有限而被统一归一化和标签化的情况将会改变。例如，职场招聘时常见的描述——“有 10 年工作经验的资深员工”，这句话本身就是前大数据时代非常粗犷的分类，符合这个条件的员工可能本身差异非常大。而在有多维度大数据的情况下，对该员工的描述不但可以更加聚焦在行业经历，也可以获得个人性格、诚信等方面的一系列信息，为企业更精准地找到合适的员工提供帮助。

当然，现在大数据维度仍然是有限的，但与区块链技术结合之后，越来越全面的个体画像显然会是未来的趋势。而与此对应，特定社会角色对候选人的精准要求也会更加明确。可以说，充分发挥每个人的个性化特点和优势，真正实现“天生我才必有用”的时代即将到来！

第二，通过大数据平台可以真正实现人工智能和人脑的平等对话，并通过这种方式互相吸取彼此的逻辑思维特点和优势，从而出现一种类人类的智慧生命体。

打破标签化定义、实现个性化身份的背后是对人机界限的突破。现代科技和大数据共同孕育出的人工智能，连接上被区块链技术优化后的个性化互联网，有可能作为独立的互联网个体参与社会和经济活动，实现人机双方以智慧生命体身份进行的平等对话。

这种通过深度学习制造出的以谷歌阿尔法狗为代表的人工智能，确实会最终创造出具有人类智慧特征的机器生命体。届时，人类的大部分劳动都将被其取代，智能生产、智能服务、智能创造将贯穿整个人类社会。

8.2.4 完善政府治理能力

数据的真实性、准确性、完备性是现代社会运行的基础，无论对政府还是工农商学兵都不可或缺。但数据又是很不可靠的，因为经手数据的是人。各种采集结果的疏漏，各种原因导致的数据不完备，各种各样的瞒报和虚假数据，都不鲜见。那么，“区块链+大数据”的组合是否可以解决从数据采集到数据质量信任的问题？

由于亚洲地理的特性，中国很早就成为一个幅员辽阔、资源丰富和人口众多的大国，拥有完整的上层建筑，而另一方面则是下层数千万乃至上亿自耕农依然处于小农经济的状态。在技术手段落后的情况下，中央只能通过官僚主义自上而下的压力，借助薄弱的社会中层直接与上亿自耕农发生关系。因此，先进的中国农耕文明在面对文明程度更低但组织动员简单、全民皆兵的北方游牧民族时，往往在军事上反而处于劣势。

同时，又由于官僚主义无法与民间商业力量产生互信，无法支撑城市文明有效发展出金融、保险、科技、物流等民间商业机构，也就无法刺激资本主义市场经济发展和科技进步。因此，国家无法与中下层人民发生高质量的联系，无法有效组织和动员全国的经济和战争潜力。这种现象在近代的鸦片战争直至抗日战争中都有体现。

到了今天，借助包括大数据、区块链以及人工智能等技术的进步，巨大国家有可能在原有的基础上进一步提高与每一个个体的联系质量，从而在拥有巨大数量的同时获得质量的提升。这是巨大国家在未来将大大优于小国的基础。

仅仅由于最近 20 年的电子政务、互联网发展，政府就已经可以收集每一

位公民的信息。当然，在初步完成各部门的数据采集后，不可避免地发展出类似数据质量低、数据孤岛、数据更新频率低等情况。但是技术本身，特别是数据采集方式的丰富和数据来源的多样性，以及区块链技术与大数据的结合，将会有效地提高中央政府和社会之间的联系。

国家不断推进完成二代身份证的普及，以及二代身份证芯片本身不断扩展的应用领域（例如，可以直接刷身份证乘坐高铁），这就是国家与每一位公民发生直接的高质量联系的案例。未来通过身份证号码可以获取的信息会包括医疗和 DNA、指纹和面部识别、个人财产和缴税情况、人员流动情况、学历情况等。在此基础上，无论是累进税推行，还是包括人口政策、扶贫攻坚等社会治理重大课题，都有了大数据背景的决策支持。在大数据这个词流行以前，中央已经可以在各省上报耕地数据之外，通过卫星图像识别和计算全国总耕地面积，保证 18 亿亩耕地红线。类似的案例还包括国家可以动用不同数据源，通过车辆保有信息有效识别贫困户信息（有车辆所有权的不是贫困户）。

而且，通过包括区块链在内的技术，政府可以精准地投入资源。例如，贵阳市政府试点的社会积分项目，通过基于区块链的底层将政府资源、社会慈善服务和困难户建立互相信任的互联，同时具备未来积分兑换和消费的可扩展性。与之类似的包括数字货币等在内的技术，可以有效提高对政府资源投入的监控、流向监测和分析。

最后，政府作为公平交易的促进者和法规监督者，将可以有效监控通过区块链互联的各方产生的具有共识的应用数据，涉及慈善互助、票据发行和交易、电子病历共享等一系列应用。这将有效促进社会资源的配置，减少政府的管理成本。

中国拥有世界上最多的人口，也必将成为世界最大的数据生产国。同时，中国拥有汇聚了众多优秀理工科人才的高等学府，有望培养出在欧美十分紧俏的数据科学家。中国可以在成为世界最大的数据生产国的同时，还具备有效运用这些数据的潜力。

大数据解决了数据来源，区块链解决了数据防伪、共享、对账勾稽等因素，加上人工智能等决策技术，将大大提升类似中国这样的巨大人口、地理、工业门类国家的整体整合水平，大大增强中央政府对全国的信息获取、资源调查、人民动员的能力。

8.3 “区块链+大数据”共创新天地

225

8.3.1 政府层面推动是前提

大数据战略得以推动的前提是充分开放的数据。美国是最早意识到这一点的国家，并通过颁布总统令等措施推动政府数据的公开。英国和法国也推出了一系列大数据开放政策。

2016年12月7日的国务院常务会议通过了《“十三五”国家信息化规划》，明确表示，信息孤岛要坚决打通，起码政府系统不应再有。会议确定的规划重点，首先便是打破信息壁垒和孤岛，构建统一高效、互联互通、安全可靠的国家数据资源体系，打通各部门信息系统，推动信息跨部门、跨层级共享共用。这说明，我国政府已经从源头做出了明确的方向性指导。

产业扶持

大数据在各行各业的推广应用离不开产业政策的扶持。大数据应用比较成熟的国家都是从政府部门开始推动的，国防部、能源部及健康管理部门一方面开放数据，另一方面成立项目来推动大数据技术的研发和应用。

在产业扶持方面，《英国数据能力战略》指出，英国政府将通过多种途径为大数据产业提供扶持；在资金支持方面，英国政府将为本国公司及有关组织提供更多机遇和便利，以获取欧盟研究与创新资金——“地平线 2020”（Horizon 2020）。

“地平线 2020”是欧盟推出的最大的一项研究与创新计划，2014 年至 2020 年 7 年间总投入资金达 800 亿欧元。英国政府此举旨在提升对欧洲科学水平的资金支持，同时将各类大数据分析中心纳入“英国资本投资战略框架”中，促进大数据分析技术的研发与产业应用。在产学研结合方面，英国还通过建立研究成果展现门户、搭建多种合作交流平台等方式，促进产业与各类研究、学术机构之间的合作和成果转化。

资金保障

资金保障是国家在推动新技术发展时常用的扶持政策之一。继美国国防部和卫生研究院宣布投资两亿多美元促进大数据研发后，英国、法国也相继宣布政府对大数据的投资。

2013 年 1 月，英国财政部明确将投入 1.89 亿英镑用于大数据和节能计算技术的研发，旨在提升地球观测和医学等领域的大数据及分析能力。同年 4 月，英国经济和社会研究委员会又宣布将新增 6400 万英镑用于大数据研发。

其中 3400 万英镑将用来建立“行政数据研究网络”，用于汇聚政府部门和机构所收集的行政数据，促进发挥政府数据对科学研究、政策制定和执行的作用。法国政府在 2013 年投入 1150 万欧元，促进法国大数据发展。

这些国家对大数据的投资体现出一定的共性特征：第一，投资领域均是关乎国家竞争力和全民生活福祉的重要领域，这些领域仅凭市场资本无法推动；第二，强化投资的核心目的是提高关键领域的大数据技术能力，它是市场化应用的前提。

强化监管

开放是原则，监管是基础。将“区块链+大数据”置于一个监管到位的环境，才可去伪存真，实现健康发展。

目前世界各国制定的监管措施多数都聚焦于数据隐私的保护层面，其中美国的数据隐私保护监管走在世界前列。美国在消费领域、健康领域都制定了一系列数据监管措施——《数据隐私保护法案》。

相比美国，其他国家的政策目前都处于指导立法的层面。英国《开放数据白皮书》明确将在公共部门透明度委员会（监督各部门数据开放的核心机构）中设立一名隐私保护专家，确保数据开放过程中及时掌握和普及最新的隐私保护措施，同时还将为各个部门配备隐私专家；二是内阁办公室强制要求所有政府部门在处理涉及个人数据时都要执行个人隐私影响评估工作（Privacy Impact Assessments），为此还专门制定了非常详细的《个人隐私影响评估手册》；三是各政府部门开放数据策略中均明确将开放数据划分为大数据（Big Data）和个人数据（My Data）。大数据是政府日常业务过程中收集到的数据，可以对所有人开放，而个人数据仅对某条数据所涉及的个人自己

开放。在数据安全方面，澳大利亚政府于 2012 年 7 月发布了《信息安全管理指导方针：整合性信息的管理》，为海量数据整合中涉及的安全风险管理提供了最佳实践指导。

总之，未来大数据将成为一切信息化战略的出口。无论“互联网+”，还是物联网、智能硬件，数据都将是价值最大的领域。同时，随着各领域数据的不断开放，在数据分享、监管方式、隐私保护¹等方面都对我们提出了更高的要求。而这些问题通过与区块链技术的深度融合，将会构建一个全新的数字信息时代。

8.3.2 企业是主体和核心

随着传感器网络的发展和智能手机的普及，数据的收集逐步自动化。特别是随着给予区块链技术的人工智能深度学习的演进，未来像生活日志、服务器日志这样的日志数据将会迎来爆炸式的增长。如何从这些庞大的乍看只是数值、文字、符号罗列的数据中挖掘“金矿”并有效运用，就必须要有广泛的企业参与。

数据管理

一般而言，数据管理主要是依据数据的价值性予以区隔，如按照招徕客

1 数据隐私的跨国合作也是亟需推动的领域。2000 年 12 月，美国商业部与欧盟签订了一份名为安全港的协议。安全港协议要求：收集个人数据的企业必须通知个人其数据被收集，并告知他们将对数据所进行的处理；企业必须得到允许才能把信息传递给第三方，必须允许个人访问被收集的数据，并保证数据的真实性和安全性，以及采取措施保证这些条款得到遵从。安全港协议确立了美国和欧盟之间隐私手续的框架，15 个成员国中签署协议的企业都将服从该协议。这意味着企业必须经个人授权后才能被第三方使用与转移。超过 4000 家与数据相关的美国企业都签署了这份协议，而未加入安全港的企业也必须单独从各个欧洲国家获取授权。

户所需的体现差异化的核心数据、除核心数据之外的背景数据这两个维度进行分类。

目前的做法是核心数据被看作企业的战略性资产，传统思路是直接保护起来，不对外提供。然而，随着区块链技术的出现，最近出现的一些案例表明，如果这些数据对自己公司有很大的好处，那么可以通过与其他公司进行战略合作的方式对数据进行共享和交换——基于 BaaS(Blockchain as a Service) 的模式。今后，我们不应只考虑保护这些数据，而更应考虑和其他公司进行分享（见图 8-4）。

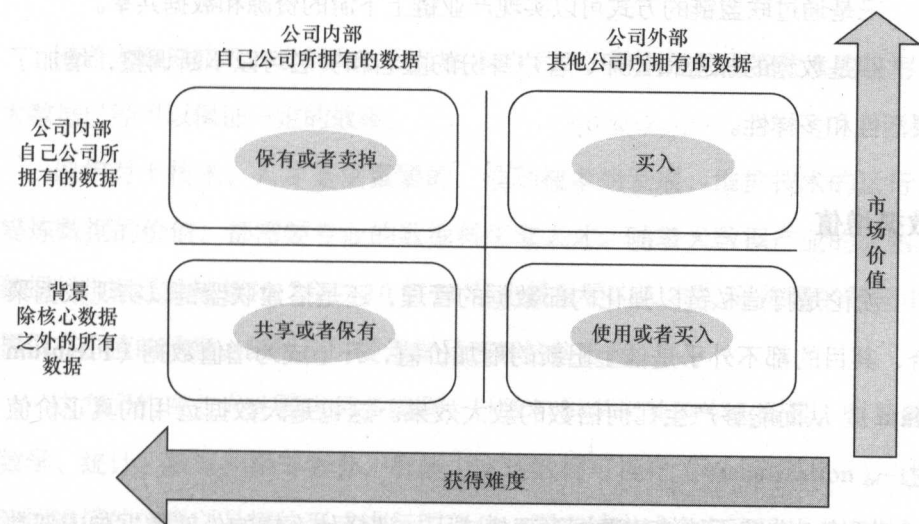


图 8-4 企业数据共享和交换决策

数据聚合

上文提到的数据出售或购买的思路很好，国家也已经认识到了数据的价值。可是现实告诉我们，有一些数据（如其他公司的顾客购买记录等）是花

钱也很难买到的。但是，要想在大数据时代确立企业的竞争优势，在战略层面必须要考虑到产业链的数据聚合。

虽然谁都有可能做到数据聚合，但在现实的中心化系统中，作为数据入口、与客户接触点最多的企业，就最有可能近水楼台先得月。而他们一旦得到了数据，就会高筑壁垒，敝帚自珍。

区块链技术可能正在改变这个局面。

一是未来对于客户隐私数据的保护将更加便捷。

二是基于公共应用的数据将会实现开放统一。

三是通过联盟链的方式可以实现产业链上下游的资源和数据共享。

四是数据的加密和公开、客户身份的匿名和实名可以不断调整，增加了灵活性和多样性。

数据增值

无论是打造私链以强化内部数据的管理，还是搭建联盟链以实现数据聚合，其目的都不外乎是催生出新的附加价值，升华成为增值数据（Premium Data），从而能够产生几何倍数的放大效果。这也是大数据运用的真正价值之一。

当然，在基于传统的数据管理模式下，选择什么样的外部数据与内部数据整合，这需要想象力。同时，在自己公司内部认为已经没什么用的数据，对于其他公司来说很可能就是求之不得的宝贝。这同样需要想象力。而通过区块链技术，这些将都不再是问题，各方只需要将可以公开的数据共享即可。

例如，耐克提供了一款面向 iPhone 的慢跑应用 Nike+GPS。它可以通过使用 GPS 在地图上记录跑步的路线，将这些数据匿名化并进行统计，就可以

找出跑步者最喜欢的路线。在体育用品店看来,这样的数据在讨论门店选址计划上是非常有效的。此外,在考虑具备淋浴、储物柜功能的收费休息区以及自动售货机的设置地点、售货品种时,这样的数据也是非常有用的。

基于区块链技术的大数据,可以让各参与方通过更加开阔的视野来发掘和制定企业战略,同时,基于技术的可视化特点,未来将培育出一批具有全局(全产业链)意识、通晓产业特点、胸怀行业战略的新一代优势企业,谁能够将区块链技术与大数据结合得更深入,谁就更有希望把握制高点。

8.3.3 热门新职业:数据科学家

随着大数据技术的发展和软硬件成本的下降,特别是区块链技术的应用,大数据已经可以保证一定的效率。

231

但相对于技术,人才是更重要的。推动技术的发展,维护技术的运行,提炼数据的价值,都需要专业的数据科学家人才。随着大数据产业的火热,数据科学家已经被誉为“今后10年IT行业最重要的人才”。从海量的数据中提取有价值的信息,最稀缺的要素就是理解这些数据的人。

这方面的专业人才要求具备编程、计算机科学相关的专业背景,要具备数学、统计、数据挖掘等素养,有能力实现数据可视化(Visualization),让晦涩难懂的信息以易懂的形式进行图形化。图存储方式会在可见的未来得到快速发展。

因为数据科学家需要具备广泛的技能和素质,因此这一岗位会成为紧缺人才,各国对人才的争夺将会出现白热化。仅仅四五年前,对国内数据科学家的需求还仅限于BAT等互联网企业中。然而最近无论是哪个行业,龙头企业都在招募基于大数据和区块链的技术团队。我们也认为,随着对区块链与

大数据结合的不断深入，未来岗位缺口将更加显现。

为了解决这个问题，笔者在《区块链世界》一书中就提到：“探索联合高校、研究机构建立区块链技术实验室，加大核心信息技术的投入，增强密码学、人工智能等方面的相关研究，跟踪量子计算机等前沿研究，增强原始创新能力，形成产学研用相结合的开发联合体，推动科研成果转化。”

如果说过去 20 年大学最热门的是经济类院系，那么我们有理由相信，下一个 10 年，数据科学将异军突起，不遑多让。未来的企业架构中还需要设立一个专门负责管理企业数据战略的 CDO(Chief Data Office，首席数据官)，他将横跨战略、市场和技术等多个维度，真正成为复合型的数字时代管理者。

让我们拭目以待这一天的到来！



分类建议：IT经济/区块链/大数据
人民邮电出版社网址：www.ptpress.com.cn

ISBN 978-7-115-45740-0



9 787115 457400 >

ISBN 978-7-115-45740-0

定价：59.00元